

The Lyman- α Forest Power Spectrum from the Sloan Digital Sky Survey

Patrick McDonald¹, Uroš Seljak¹, Scott Burles², David J. Schlegel³, David H. Weinberg⁴, David Shih¹, Joop Schaye⁵, Donald P. Schneider⁶, J. Brinkmann⁷, Robert J. Brunner⁸, Masataka Fukugita⁹

ABSTRACT

We measure the power spectrum, $P_F(k, z)$, of the transmitted flux in the Ly α forest using 3035 high redshift quasar spectra from the Sloan Digital Sky Survey. This sample is almost two orders of magnitude larger than any previously available data set, yielding statistical errors of $\sim 0.6\%$ and ~ 0.005 on, respectively, the overall amplitude and logarithmic slope of $P_F(k, z)$. This unprecedented statistical power requires a correspondingly careful analysis of the data and of possible systematic contaminations in it. For this purpose we reanalyze the raw spectra to make use of information not preserved by the standard pipeline. We investigate the details of the noise in the data, resolution of the spectrograph, sky subtraction, quasar continuum, and metal absorption. We find that background sources such as metals contribute significantly to the total power and have to be subtracted properly. We also find clear evidence for SiIII correlations with the Ly α forest and suggest a simple model to account for this contribution to the power. While it is likely that our newly developed analysis technique does not eliminate all systematic errors in the $P_F(k, z)$ measurement below the level of the statistical errors, our tests indicate that any residual systematics in the analysis are unlikely to affect the inference of cosmological parameters from $P_F(k, z)$. These results should provide an essential ingredient for all future attempts to constrain modeling of structure formation, cosmological parameters, and theories for the origin of primordial fluctuations.

¹Physics Department, Princeton University, Princeton, NJ 08544, USA; pm,useljak@princeton.edu

²Physics Department, MIT, 77 Massachusetts Av., Cambridge MA 02139, USA

³Princeton University Observatory, Princeton, NJ 08544, USA

⁴Department of Astronomy, Ohio State University, Columbus, OH 43210, USA

⁵School of Natural Sciences, Institute for Advanced Study, Einstein Drive, Princeton NJ 08540, USA

⁶Department of Astronomy and Astrophysics, The Pennsylvania State University, University Park, PA 16802, USA

⁷Apache Point Observatory, 2001 Apache Point Rd, Sunspot, NM 88349-0059, USA

⁸NCSA and Department of Astronomy, University of Illinois, Urbana, IL 61801, USA

⁹Inst. for Cosmic Ray Research, Univ. of Tokyo, Kashiwa 277-8582, Japan

Subject headings: cosmology: data analysis/observations—intergalactic medium—large-scale structure of universe—quasars: absorption lines

1. Introduction

Although the Ly α forest was discovered many decades ago (Lynds 1971), it has only recently emerged as one of the prime tracers of the large scale structure in the Universe. The high resolution measurements using the Keck HIRES spectrograph (Vogt et al. 1994) have been largely reproduced using hydrodynamical simulations (Cen et al. 1994; Zhang et al. 1995; Hernquist et al. 1996; Theuns et al. 1998) and semi-analytical models (Gnedin & Hui 1998). The picture that has emerged from these studies is one in which the neutral gas responsible for the absorption is in a relatively low density, smooth environment, which implies a simple connection between the gas and the underlying dark matter. The neutral fraction of the gas is determined by the interplay between the recombination rate (which depends on the temperature of the gas) and ionization caused by ultraviolet photons. Photoionization heating and expansion cooling cause the gas density and temperature to be tightly related, except where mild shocks heat up the gas. This leads to a tight relation between the absorption and the gas density. Finally, the gas density is closely related to the dark matter density on large scales, while on small scales the effects of thermal broadening and Jeans smoothing must be included. In the simplest picture described here, all of the physics ingredients are known and can be modeled. The fact that one can trace the fluctuations over a range of redshifts ($2 \lesssim z \lesssim 6$ using ground based spectrographs) and over a range of scales, which are typically smaller than the scales of other tracers, is the main strength of this method. It becomes particularly powerful when combined with cosmic microwave background (CMB) anisotropies or other tracers that are sensitive to larger scales. Such a combination is sensitive to the shape of the primordial spectrum of fluctuations, which is one of the few observationally accessible probes of the early universe. These observations are therefore directly testing the models of the early universe such as inflation.

Ly α forest observations and constraints on cosmology have been explored by several groups in the past. Most of the analyses focused on the power spectrum, $P_F(k)$, of the fluctuations in the Ly α forest flux,

$$\delta_F(\lambda) = \exp[-\tau(\lambda)] / \langle \exp(-\tau) \rangle - 1, \quad (1)$$

where τ is the optical depth to Ly α absorption. The first such work was by Croft et al. (1998), followed by McDonald et al. (2000), Croft et al. (2002), and Kim et al. (2003). These groups were limited to a few dozen spectra at most. Recent theoretical analyses, in addition to above, have been performed by Gnedin & Hamilton (2002), Zaldarriaga et al. (2001), and Seljak et al. (2003). In the latter two of these papers the degeneracy between the amplitude and slope of the primordial power spectrum and the normalization of the optical depth-density relation [most sensitive to the

intensity of UV background, and typically parametrized in terms of the mean transmitted flux fraction, $\bar{F} \equiv \langle \exp(-\tau) \rangle$ was emphasized, which leads to a significant expansion of the allowed range of cosmological parameters relative to what one would have inferred from the errors on the flux power spectrum alone. Seljak et al. (2003) have shown that the current Ly α forest constraints are consistent with the Λ CDM model favored by recent CMB data, testing it in a regime of redshift and length scale not probed by other measurements, but that within the Λ CDM framework they do not add much leverage on parameter values beyond that afforded by the CMB data alone.

An important practical implication of the theoretical breakthroughs of the 1990s is that large scale structure in the Ly α forest can be effectively studied with moderate resolution spectra. Once the spectrum is modeled as a continuous phenomenon rather than a collection of discrete lines, there is no need to resolve every feature. Some of the studies cited above use high resolution ($\sim 0.08\text{\AA}$) spectra, some use moderate resolution ($\sim 1 - 3\text{\AA}$) spectra, and some use a mix of the two.

The goal of this paper is to present a new measurement of the Ly α forest power spectrum, based on ~ 3000 Sloan Digital Sky Survey (SDSS; York et al. (2000)) spectra that probe the Ly α forest at a resolution $R \sim 2000$ ($\sim 2.5\text{\AA}$ FWHM). This sample is almost two orders of magnitude larger than anything that was available before. As such it greatly increases the statistical power of the Ly α forest, making it comparable to the CMB from WMAP. At the same time, the required tolerance of systematic errors also increases by the same amount. This requires a careful investigation of all of the sources of systematic errors, and a large portion of this paper is devoted to the issue of possible systematics in the data and their influence on the parameters of interest. We also discuss how the analysis we perform and results we obtain differ from what can be done using the standard spectral pipeline outputs in the public SDSS data. In part because of the practicalities of work in a large, multi-institutional collaboration, and in part because of the importance of obtaining an accurate measurement with well understood statistical and systematic errors, the Ly α forest power spectrum has been pursued by two independent groups within the SDSS, one led by P. McDonald and U. Seljak, and the other by L. Hui and A. Lidz. The methods employed are different and have been developed independently. Results of the alternative analysis will be presented elsewhere (Hui et al., in preparation).

We only present the observational measurement of the SDSS Ly α forest power spectrum in the current paper. Independent of any theoretical interpretation, this basic result should be robust on the scales for which we give results, $0.0013 (\text{km s}^{-1})^{-1} < k < 0.02 (\text{km s}^{-1})^{-1}$, where $k \equiv 2\pi/\lambda$ if λ is the wavelength of a Fourier mode (not to be confused with spectral wavelength), here measured in km s^{-1} (note that throughout the paper we frequently use velocity in place of observed wavelength, with the understanding that all that enters into our calculations are velocity differences between pixels of measured spectra, defined by $\Delta v = c \Delta \ln(\lambda)$ – we do not measure power on scales large enough for the imperfections in this expression to become relevant). The choice of k -range is determined by the continuum fluctuations on the low end and spectral resolution at the high end. We note also that the useful range is limited not only by these uncertainties, which are related to the data analysis, but also by the uncertainties in the theoretical modeling and/or additional

astrophysical effects. We will address these latter issues in more detail in a separate publication. However, we do not completely decouple the theory from the data analysis. For example, when discussing the importance of systematic errors it is useful to understand how they would affect cosmological results like the slope and amplitude of the matter power spectrum, so much of our discussion of systematics is devoted to this issue.

The common usage of the term Ly α forest is to describe the Ly α absorption by neutral hydrogen in the relatively low density bulk of the IGM. In this paper we include damped-Ly α systems (DLAs) in the definition of the “forest”, so it includes all HI-Ly α absorption. We could try to remove DLAs before measuring $P_F(k, z)$, because they are more difficult to simulate than the lower optical depth absorption; however, we believe the advantage of removal is illusory. If the DLAs were located randomly within the IGM (which they certainly are not completely), it would be simple to include them in the theory using their known column density distribution. If they are not located randomly, the regions obscured by DLAs in the spectra are special, so the effect of removing the DLAs still must be understood using simulations. We leave the handling of the effects of DLAs as a problem for the theory, which we will address elsewhere.

Absorption by metals is also difficult to simulate accurately, so we would like to remove this contribution to $P_F(k, z)$. This is relatively easy to do for transitions with wavelength $\lambda \gtrsim 1300$ Å, but it is basically impossible for transitions with $\lambda \lesssim \lambda_\alpha$, because the metal features always appear mixed with HI-Ly α . We will subtract the power measured in the rest wavelength range $1268 \text{ Å} < \lambda_{\text{rest}} < 1380 \text{ Å}$ from our measurement of the power in the forest, which removes the effect of transitions with longer wavelength, but we leave shorter wavelength transitions as part of the forest. The only significant contaminant of this kind that we can identify is SiIII absorption at 1206.50 Å , and we develop a simple and effective way to account for this in the theory. We refer to our final background-subtracted power spectrum as $P_F(k, z)$, and use $P_{\lambda_1, \lambda_2}(k, z)$ for the raw power measured in the interval $\lambda_1 < \lambda_{\text{rest}} < \lambda_2$. We are using the range $1041 \text{ Å} < \lambda_{\text{rest}} < 1185 \text{ Å}$ for the Ly α forest.

The outline of this paper is as follows. In §2, we describe the selection of our data set and the preparation of spectra for the measurement of $P_F(k, z)$. In §3 we describe the method used to measure the power spectrum and estimate the error bars, test the procedure, and give the basic results. We perform consistency checks on the results and discuss systematic errors in §4, which is followed by a brief recipe for using our results in §5, and conclusions in §6.

2. Data Selection and Preparation

We describe the sample of quasar spectra that we use in §2.1. In §2.2 we explain how we remove broad absorption line (BAL) quasars from the sample. In §2.3 we explain how we combine spectra from different exposures for the same quasar and use the differences between exposures to understand the noise in the data. We discuss the resolution of the spectra in §2.4. Finally, in §2.5

we describe how we divide each spectrum by an estimate of the quasar continuum, the expected mean absorption level in the spectrum, and a spectral calibration vector (see below), to produce the vectors of transmission-fluctuation estimates, δ_f , for each quasar, from which we will measure $P_F(k, z)$.

2.1. SDSS Observations and Sample Selection

The Sloan Digital Sky Survey (York et al. 2000) uses a drift-scanning imaging camera (Gunn et al. 1998) and a 640 fiber double spectrograph on a dedicated 2.5 m telescope. It is an ongoing survey to image 10,000 sq. deg. of the sky in the SDSS *ugriz* AB magnitude system (Fukugita et al. 1996; Stoughton et al. 2002) and to obtain spectra for $\sim 10^6$ galaxies and $\sim 10^5$ quasars. The astrometric calibration is good to better than $0''.1$ rms per coordinate (Pier et al. 2003), and the photometric calibration is accurate to 3% or better (Hogg et al. 2001; Smith et al. 2002). The data sample used in this paper was compiled in Summer 2002 and is a combination of data releases one (Abazajian et al. 2003) and two (Abazajian et al. 2004).

About 13% of the spectroscopic survey targets are quasar candidates selected based on their colors (Richards et al. 2002). The magnitude limit for UV-excess objects is $i = 19.1$, while additional high-redshift candidates ($z > 3$) are targeted to $i = 20.2$. Fibers are allocated according to a tiling algorithm (Blanton et al. 2003), with the galaxy sample and the quasar sample being the top priorities. The remaining 8% of fibers serve for calibration purposes.

SDSS spectra are obtained using plates holding 640 fibers, each of which subtends $3''$ on the sky; the spectra cover $3800 - 9200\text{\AA}$. The pixel width is a slowly varying function of wavelength, but is typically $\sim 70\text{ km s}^{-1}$. The resolution also varies, but is typically also $\sim 70\text{ km s}^{-1}$ rms (i.e., the resolution is $1800 < R < 2100$ and there are ~ 2.4 pixels per FWHM resolution element). All quasars have multiple spectra, usually taken one after the other (timescales of a fraction of an hour), so the quasar variability can be ignored (in the opposite case it would act as an additional source of noise). The co-added spectra in the official SDSS release use local spline interpolation onto a uniform grid of pixels of width $\Delta \log_{10}(\lambda) = 0.0001$, and do not guarantee the noise to be uncorrelated. We therefore redo this step starting from the individual exposures. This is discussed in more detail below. Spectral flux errors per pixel in most cases are about $1 \times 10^{-17}\text{ erg s}^{-1}\text{ cm}^{-2}\text{ \AA}^{-1}$. Redshifts are automatically assigned by the SDSS spectral classification algorithm, which is based on χ^2 fitting of templates to each spectrum (Schlegel et al., in preparation).

We limit ourselves to quasars with redshift $z_q > 2.3$ when measuring the power in the Ly α forest region of spectra, so that each spectrum contains a significant stretch of the Ly α forest above the detector cutoff at 3800 \AA (which corresponds to Ly α absorption at $z = 2.12$). We use the sample compiled in Summer 2002, cut down to 3035 spectra by eliminating some plates of questionable quality, some spectra where two different redshift estimation codes disagree, and some BAL quasars (see below). Figure 1 shows the redshift distribution of the data. The dashed, red histogram shows

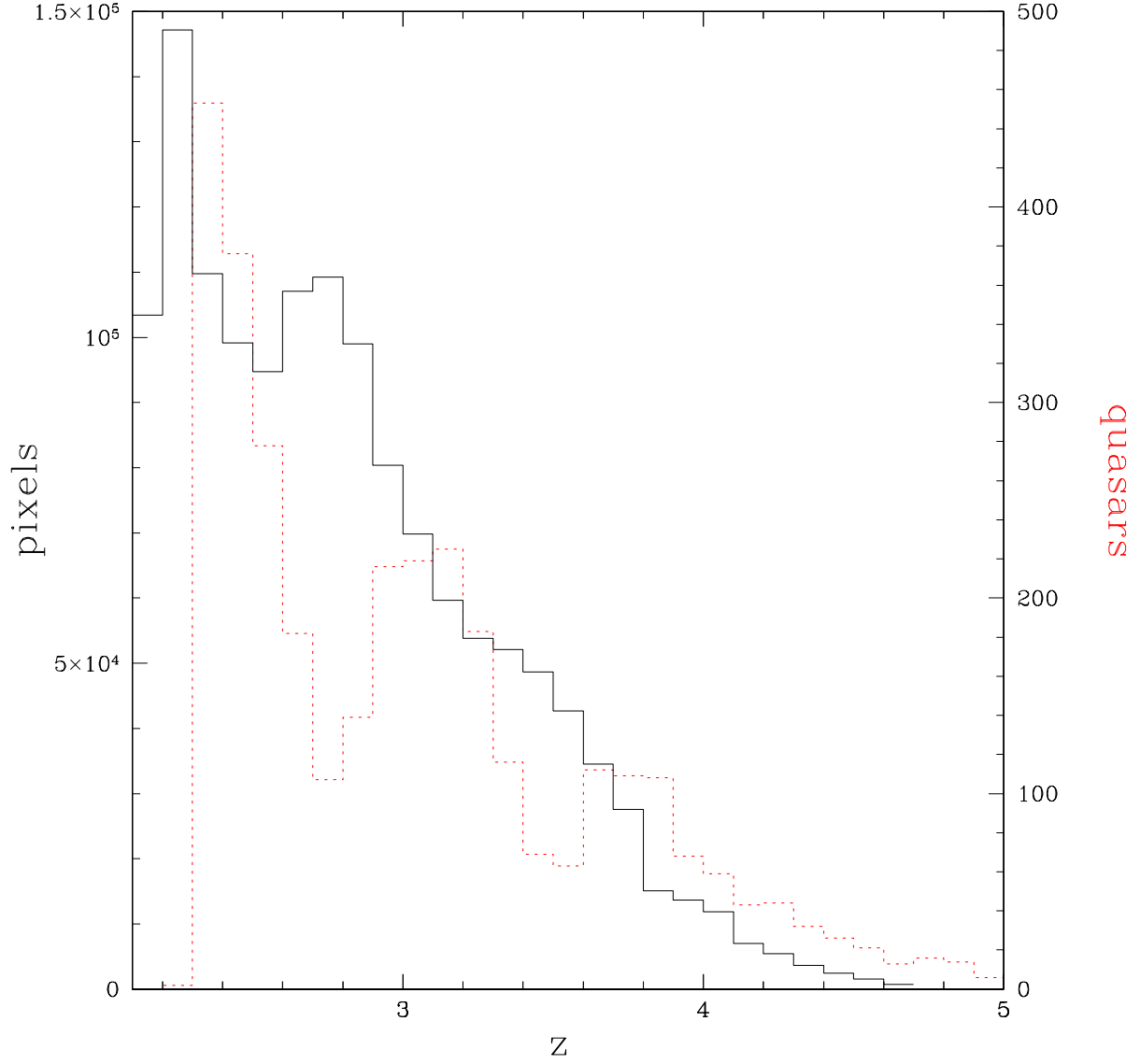


Fig. 1.— The distribution of the spectral pixels used to probe the Ly α forest (black, solid histogram; scale on left axis), and the redshift distribution of our primary sample of 3035 quasars (red, dotted histogram; right axis). Note the gap at $z \sim 2.7$ in the quasar redshift distribution, caused by a class of stars being indistinguishable from quasars in the SDSS photometry (Richards et al. 2002).

the distribution of quasar redshifts. The solid, black histogram shows the distribution of pixels in the range $1041 < \lambda_{\text{rest}} < 1185 \text{ \AA}$. Note that there is a gap in the quasar redshifts around $z \sim 2.7$, which is due to the stellar locus crossing the quasar locus in the 5-color SDSS photometry (Richards et al. 2002). Figure 2 shows an example SDSS spectrum of a $z = 3.7$ quasar. This spectrum is unusual in that most have lower S/N, and most quasars are at lower redshift.

We employ an additional sample of ~ 8000 spectra with $z_q < 2.3$, so that we can study the full observed wavelength range, $3800 \text{ \AA} \lesssim \lambda \lesssim 9200 \text{ \AA}$, outside the confusion of the Ly α forest. As we discuss in §3.4, we compute a non-negligible background power term (probably mostly metal absorption), by measuring the power in the wavelength range $1268 \text{ \AA} < \lambda_{\text{rest}} < 1380 \text{ \AA}$. Using only the primary sample, we would not be able to compute this term for observed wavelengths below $\sim 4400 \text{ \AA}$.

We remove several wavelength regions from our analysis because of calibration problems: $\lambda < 3800 \text{ \AA}$, $5575 \text{ \AA} < \lambda < 5583 \text{ \AA}$, $5888 \text{ \AA} < \lambda < 5894 \text{ \AA}$, $6296 \text{ \AA} < \lambda < 6308 \text{ \AA}$, and $6862 \text{ \AA} < \lambda < 6871 \text{ \AA}$ (the last two have no direct effect on the results we present). Most of these problems are due to strong sky lines.

2.2. BAL Removal

Our sample was initially examined by eye, and the most extreme broad absorption line (BAL) quasars were removed (see Hall et al. (2002) for a discussion of BALs). When we first measured the background power in the region $1409 < \lambda_{\text{rest}} < 1523 \text{ \AA}$, we found that the most extreme outliers in power were still obvious BALs (this was not true of the Ly α forest region). To test the importance of these systems to our Ly α forest power measurement, we removed a further 147 quasars using the following automated method: Each spectrum is smoothed by a Gaussian with rms width 280 km s^{-1} . The continuum within the region $1420 < \lambda_{\text{rest}} < 1535 \text{ \AA}$ is redefined by dividing by the mean flux-to-continuum ratio in the region. A quasar is identified as a BAL quasar if the region $1420 < \lambda_{\text{rest}} < 1535 \text{ \AA}$ contains a 2000 km s^{-1} long continuous set of pixels that all fall more than 20% above or below our estimated continuum (we initially identified wide regions with flux above the continuum out of simple curiosity, but found that these are in practice almost always obvious BAL quasars where the continuum has been biased low by the BAL feature). We iterate the continuum redefinition twice, computing the new mean after throwing out pixels more than 20% below the previous mean, but this makes almost no difference to the results. Note that the 280 km s^{-1} smoothing was applied to allow easier identification of BALs in noisy spectra. As we show below, this BAL cut makes essentially no difference to our $P_F(k, z)$ result, although it does have a noticeable effect on the power measurement in the region $1409 < \lambda_{\text{rest}} < 1523 \text{ \AA}$.

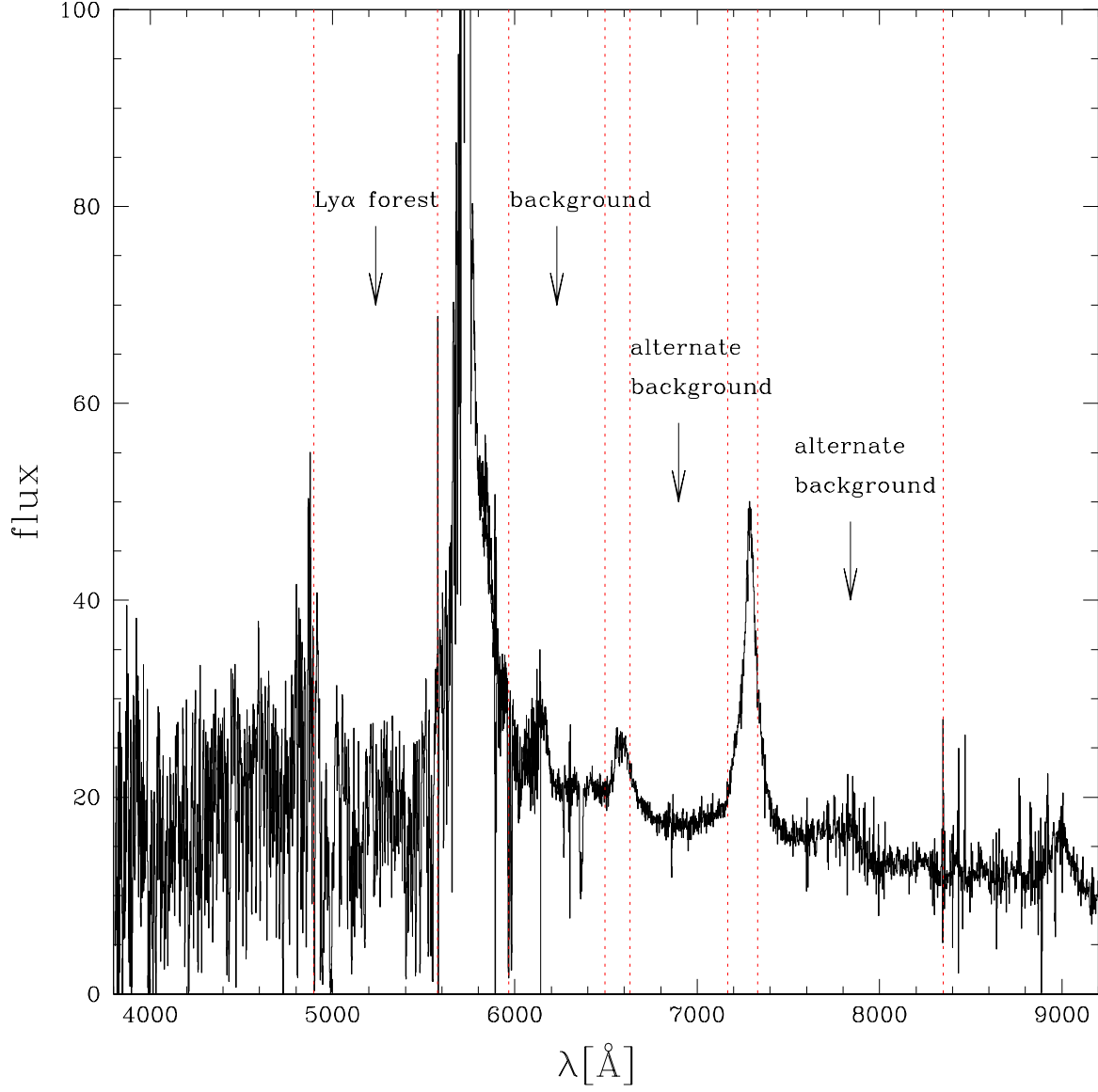


Fig. 2.— Example spectrum of a $z = 3.7$ quasar with unusually high S/N. The regions we use to measure the Ly α forest power and background power are indicated by vertical dotted lines, along with a couple of alternate regions that we will discuss (note that the background and Ly α forest observed in the same quasar spectrum correspond to different redshifts).

2.3. Combining Exposures and Calibrating the Noise

SDSS obtains multiple (at least 3) exposures for each quasar. We combine the individual exposure spectra to produce a single spectrum, using a nearest-grid-point method that produces uncorrelated noise and a reasonably well-defined sampling window. For each pixel we record estimates of wavelength, quasar flux, resolution, sky flux, read-noise, and two different total noise estimates. The first noise estimate, which we will call simply σ_p (p for pipeline), is computed using the error array given for the exposure spectra by the spectral reduction pipeline. The second noise estimate, which we call σ_c (c for component), is computed by summing the read-noise and the noise implied by estimates of the number of photons corresponding to the quasar flux and sky flux. The two noise estimates generally do not agree, but this is not a problem for us because we ultimately recalibrate the noise (next). Finally, we record χ^2/ν for each pixel, computed by treating the determination of the combined flux value for each pixel as a one parameter fit to the measurements given by the different exposures. Examples of the more important of these quantities in Ly α forest regions are shown in Figure 3. For comparison to the sky and quasar flux levels, we have converted the Gaussian read-noise into the flux of photons that would contribute the same noise variance. Several elements of Figure 3 (e.g., the estimation of the quasar continuum and σ_w) will be described later in this paper.

The noise estimate from the standard SDSS pipeline is only approximate. The accuracy of the noise estimate required for our purpose is much higher than anticipated when the pipeline was developed. For this reason we use the differences between single-exposure spectra for the same quasar to determine the noise properties of the data. We construct difference spectra by combining the flux-calibrated exposures with alternating sign for each exposure, i.e., we use exactly the same procedure that we normally use to produce combined spectra from the exposures, except half of the exposures are subtracted instead of added, so the mean result is zero (we drop the last exposure when there are an odd number – this is not the most efficient method possible, but we do not need it to be). The result is a direct measure of the exposure-to-exposure changes. We measure the power spectrum of these difference spectra using the method described in §3.1, including noise subtraction based on the pipeline noise estimates for the pixels. The result is shown in Figure 4 (points with error bars). We obtain a clear detection of power, where there should be none if the spectra differ only by the noise estimate from the pipeline which is being subtracted. If we assume that the noise has been underestimated by a constant factor, and fit for that factor using the error covariance matrix estimated by bootstrap resampling, we find a decent fit: $\chi^2 = 141.6$ for 107 degrees of freedom (formally, this fit is not good because χ^2 is unlikely to be this high by chance). This fit uses our usual points in $0.0013 \text{ (km s}^{-1}\text{)}^{-1} < k < 0.02 \text{ (km s}^{-1}\text{)}^{-1}$. The best fit value of the excess noise contribution is $16.1 \pm 0.4\%$ of the original noise estimate, indicating that the rms noise was underestimated by 8%. The best fit and goodness of fit do not change if we add points on larger scales. The quality of the fit begins to degrade as we add points with larger k , but the best fit value changes by only 1% (in power) out to the Nyquist frequency of the data. Of course, we have no reason to expect a single redshift-independent factor to describe the relation between the true

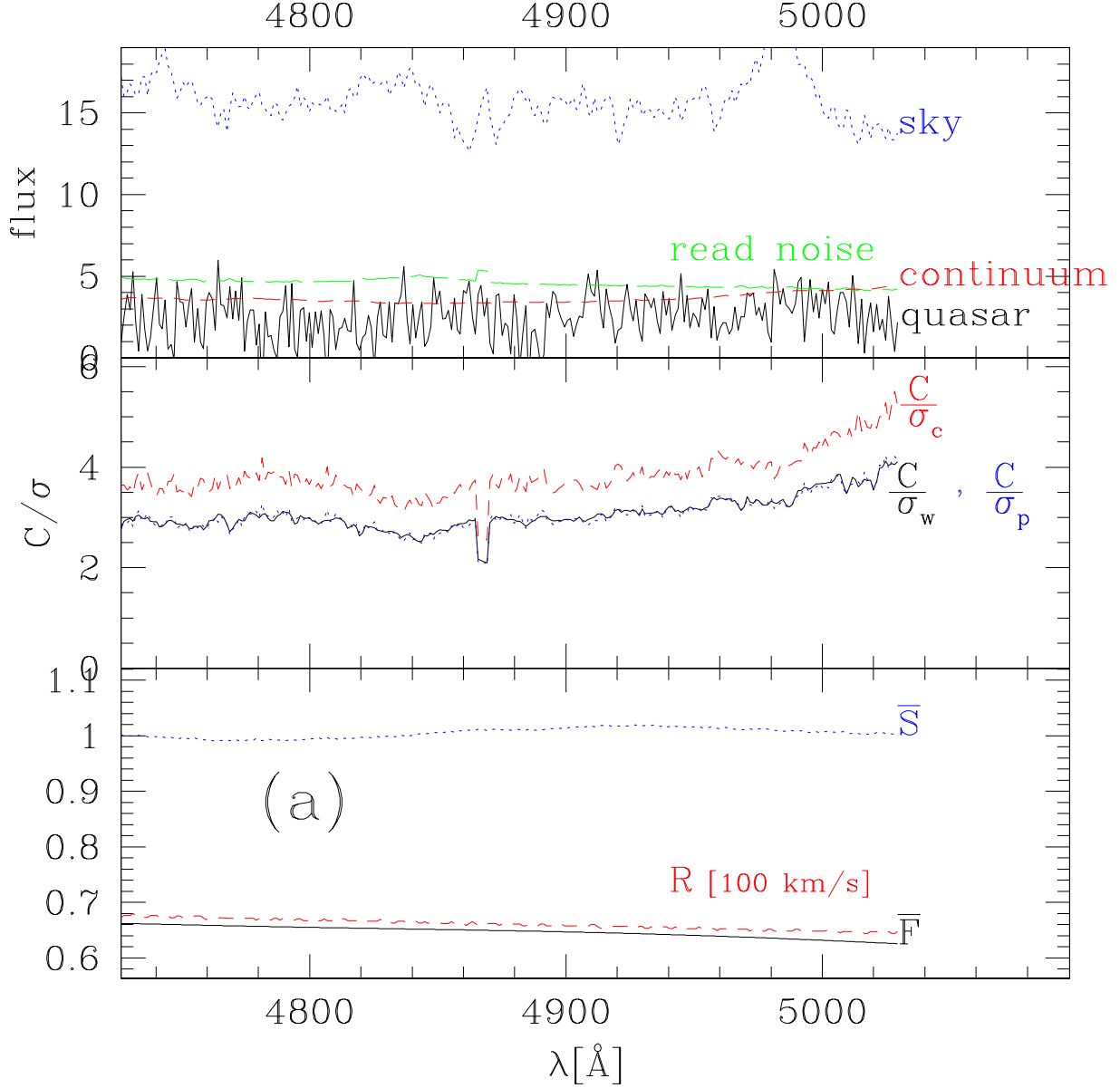
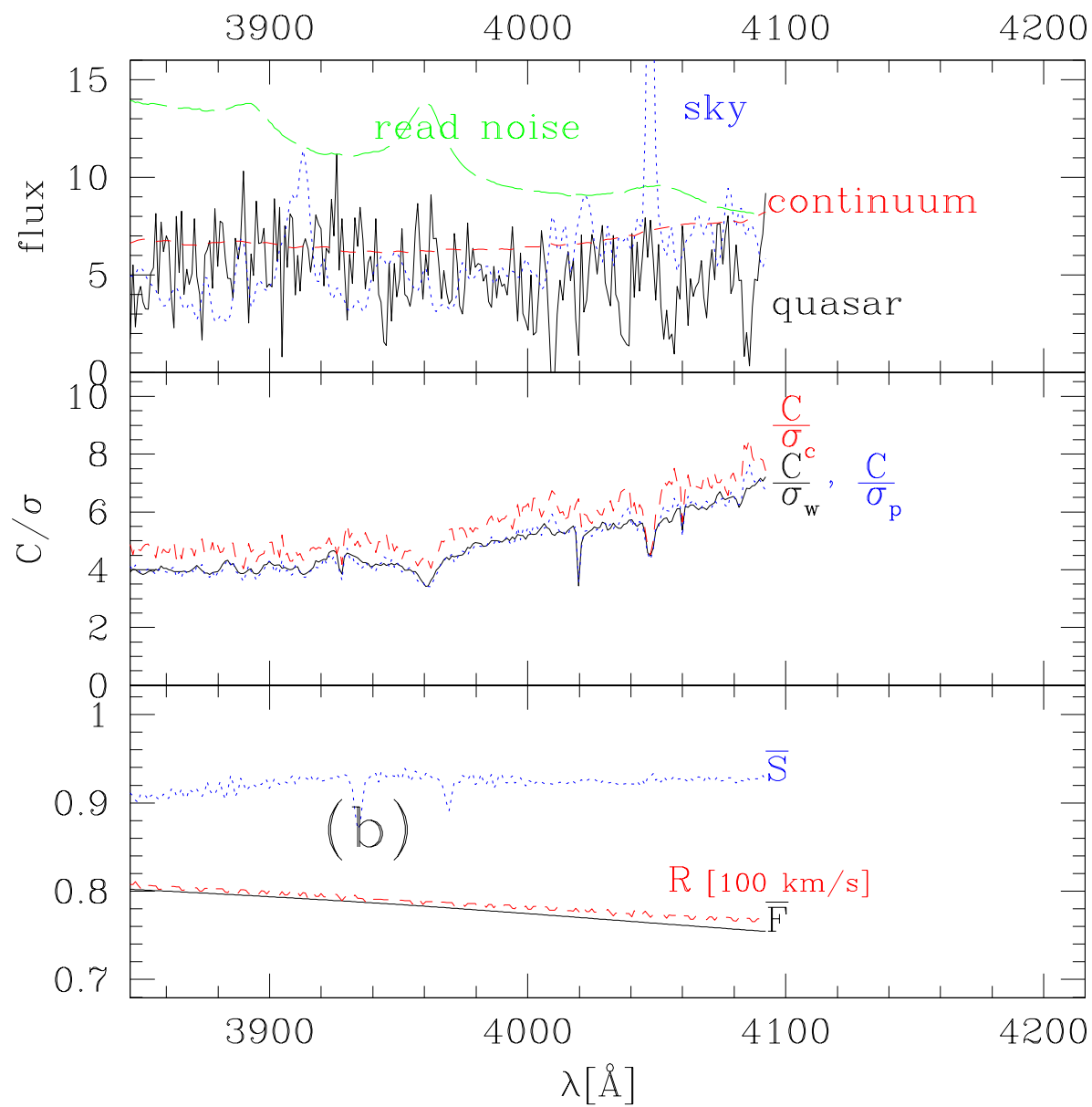
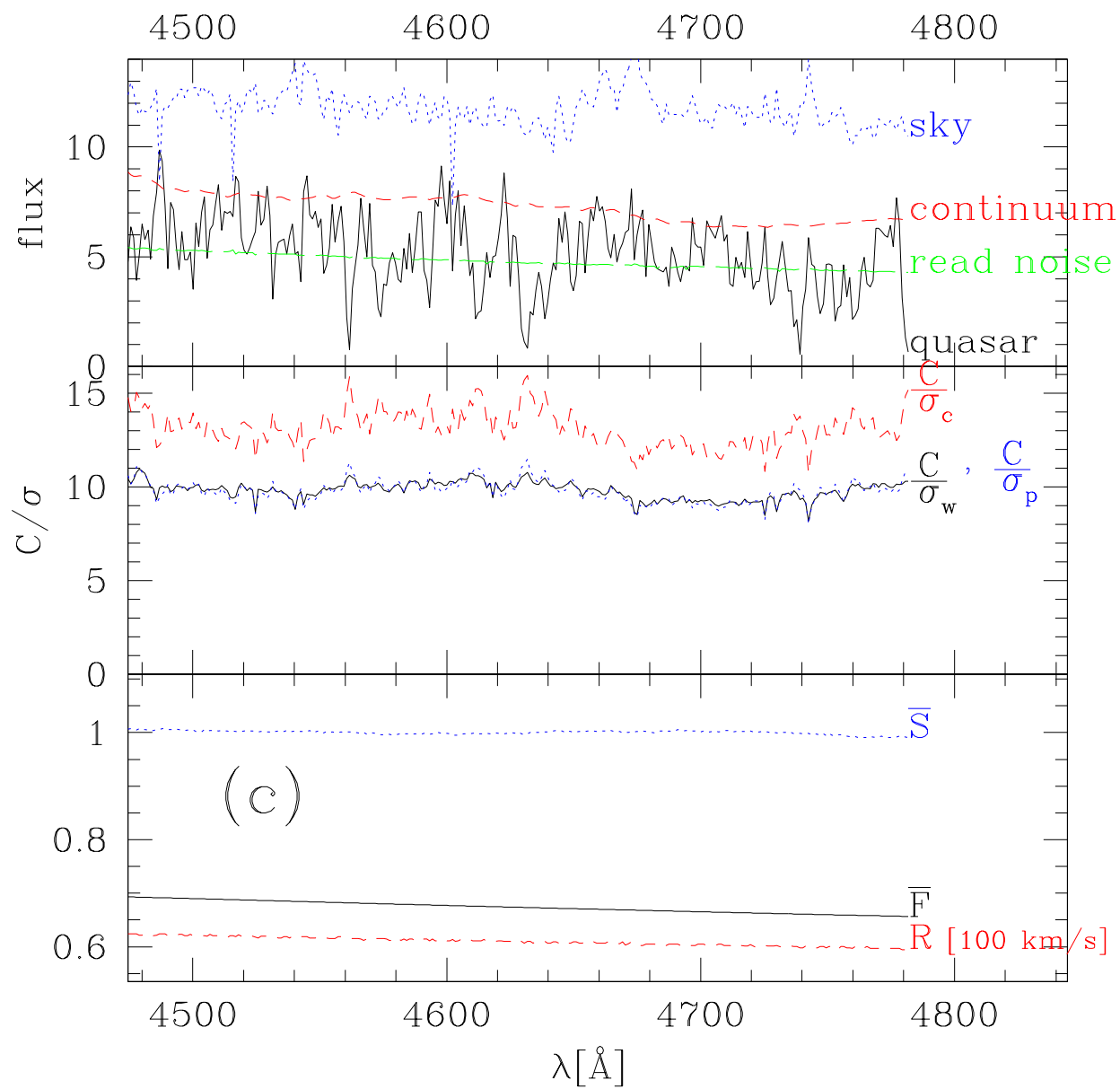


Fig. 3.— Examples of the chunks of spectra used to measure power, with (a,b) showing quasars at $z_q = (3.24, 2.45)$ over the rest wavelength range $1113 \text{ \AA} < \lambda_{\text{rest}} < 1185 \text{ \AA}$, and (c) showing a quasar at $z_q = 3.30$ over the rest wavelength range $1041 \text{ \AA} < \lambda_{\text{rest}} < 1113 \text{ \AA}$. Top panel: quasar flux (solid black line), sky flux (dotted blue line), our continuum estimate (red short-dashed line), and the read-noise as an equivalent photon flux (green long-dashed line). Middle panel: S/N level shown as a ratio of our continuum to the different rms noise levels (see text), σ_w (black solid line), σ_p , (blue dotted line), and σ_c (red dashed line). Bottom panel: Calibration correction vector, \bar{S} (blue dotted line), rms resolution in units of 100 km/s (red dashed line), and evolution of the mean transmission fraction, $\bar{F}(z)$ (black solid line). The perfect degeneracy in our analysis between the overall normalization of the continuum and $\bar{F}(z)$ has been broken arbitrarily, so only the evolution of $\bar{F}(z)$ is meaningful (see text).





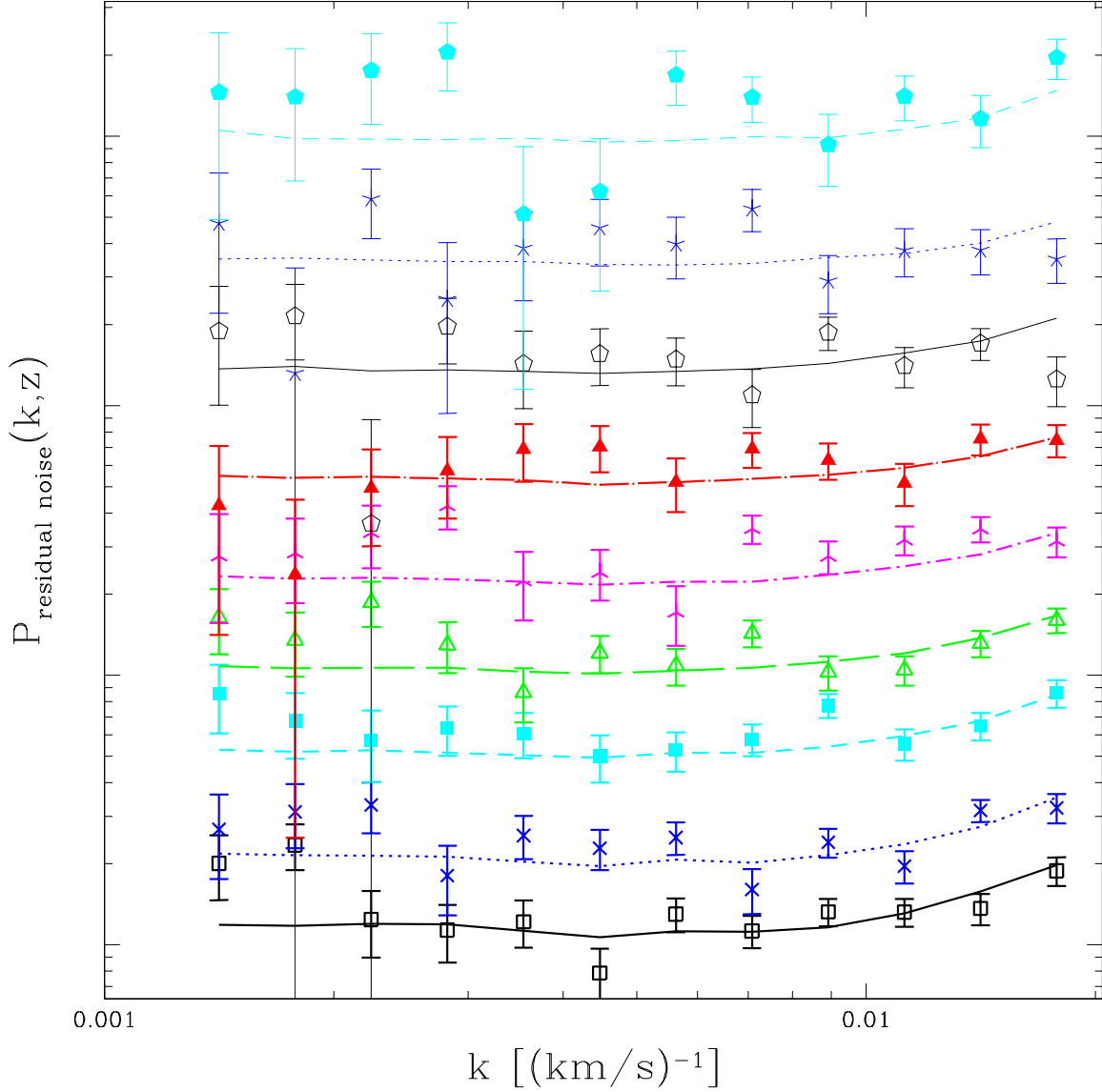


Fig. 4.— The points show the measured power in difference spectra, created by subtracting separate exposures for the same quasar. Noise power has been subtracted based on the standard pipeline noise estimates for each exposure. The lines show 16% of the subtracted noise term. The different colors, lines, and symbols identify redshift bins, in a pattern that we will use repeatedly throughout the paper. From bottom to top — $z=2.2$: black, solid line, open square; $z=2.4$: blue, dotted line, 4-point star (cross); $z=2.6$: cyan, dashed line, filled square; $z=2.8$: green, long-dashed line, open triangle; $z=3.0$: magenta, dot-dashed line, 3-point star; $z=3.2$: red, dot-long-dashed line, filled triangle; $z=3.4$: black, thin solid line, open pentagon; $z=3.6$: blue, thin dotted line, 5-point star; $z=3.8$: cyan, thin dashed line, filled pentagon. The different redshifts have been shifted vertically by arbitrary amounts on this logarithmic plot.

and pipeline noise, so the formally bad χ^2 is not a fundamental problem. We check for systematic change with redshift by allowing a power law dependence, $P_{\text{residual noise}} \propto [3.75/(1+z)]^d$, but find no significant detection ($d = 0.07 \pm 0.20$). Our final method will effectively account for evolution anyway, as described below.

A k dependence different than expected for white noise could be a problem for us, so we check for this by fitting for a power law dependence, $P_{\text{residual noise}} \propto (k/k_p)^b$ [with $k_p = 0.0074 \text{ (km s}^{-1}\text{)}^{-1}$], finding $b = -0.111 \pm 0.025$, a significant detection (χ^2 is now a reasonable 123.3 for 106 dof). Allowing a running of the power law, $P_{\text{residual noise}} \propto (k/k_p)^{b+1/2 \text{ } c \text{ } \ln(k/k_p)}$, does not improve the fit ($c = -0.046 \pm 0.066$). The slope we find corresponds to a $\sim 20\%$ change in 16% of the noise power at the extremes of our k range, i.e., only $\sim 3\%$ of the total noise power, which is a relatively small fraction of the Ly α forest power except at the highest k (see Figure 11 below). We henceforth assume that the extra noise is proportional to $k^{-0.111}$ rather than white (this makes $< 1\%$ difference in the final results except for the one highest k , lowest z point where the difference is 2%).

How accurate is this noise estimate based on differences between exposures? Our difference spectra will contain a component of the Ly α forest power if the calibration between exposures is not perfect. The power in this term would be suppressed relative to the Ly α forest power by the fractional calibration error squared, so it would be very small unless the exposure-to-exposure calibration errors were quite large. The fact that a simple one parameter extra-noise model fits reasonably well, in the face of variation in redshift, noise amplitude, and k , argues against calibration errors being a big problem. More convincingly, we measure nearly the same excess noise contribution ($14.2 \pm 0.5\%$) and slope ($b = -0.135 \pm 0.028$) in the region $1268 \text{ \AA} < \lambda_{\text{rest}} < 1380 \text{ \AA}$ as we do in the Ly α forest. This argues against any connection to leaking Ly α forest power. Note that the effective absolute level of noise in the $1268 \text{ \AA} < \lambda_{\text{rest}} < 1380 \text{ \AA}$ region is about half that in the Ly α forest region, so this test shows that the fraction of extra noise does not depend strongly on the noise level itself.

Pixels in different exposures are not perfectly aligned, and misalignment can allow Ly α forest power to leak into our difference spectra. To test this alternative explanation for the apparent excess noise in the spectra, we split the spectra into two groups with approximately equal weight, based on the rms misalignment in the forest region (the alignment is known from the wavelength calibration of the exposures, which is thought to be practically perfect). We find the same excess noise power in both the poorly aligned group ($16.1 \pm 0.6\%$, $b = -0.086 \pm 0.036$) and the better aligned group ($15.3 \pm 0.6\%$, $b = -0.123 \pm 0.036$), suggesting that the excess power is not due to misalignment. Furthermore, the presence of a similar level of excess noise power outside of the forest region again argues against leakage. We therefore believe that our noise estimate is considerably more accurate than the noise estimate from the SDSS pipeline.

In our initial power spectrum analysis we multiplied the noise-power estimated from the pipeline errors by the factor 1.16 for all spectra; however, when we split the data based on the mean value of χ^2/ν for the exposure combination (see §4.4) we found that the large and small

χ^2/ν subsamples disagreed significantly on the $P_F(k, z)$ results. We eliminated this problem by estimating the noise-correction factor individually for each spectrum, by fitting to the power in the difference spectrum for that quasar. The mean extra power from these fits is still close to 16%, but there is considerable scatter. When we use these individual estimates, the correlation between measured $P_F(k, z)$ and χ^2/ν disappears, i.e, the mean value of χ^2/ν for a spectrum’s exposure combination was a good indicator of the amount by which the noise in each spectrum was misestimated. Note that there are statistical errors in these noise estimates for each spectrum, of the same order as the error for which we are trying to correct; however, there is no systematic bias associated with these errors, and the random error they contribute is automatically included in our final bootstrap errors. In fact, including the spectrum-by-spectrum noise estimate reduces the bootstrap errors slightly on small scales, verifying that these estimates are on average more accurate than the original noise estimates. It is not known why the noise is misestimated by the standard pipeline. Tests at this level have not been done before.

Our final data product will be a measurement of $P_F(k, z)$ binned in k and z , i.e., a matrix $P_{F,ij}$ where i labels bins with z_i and j labels bins with k_j . We will also give the noise power that was subtracted, $P_{N,ij}$, in the same bins. We suggest allowing a 5% rms freedom in the noise amplitude in each z bin when performing model fits, i.e., for each bin subtract $f_i P_{N,ij}$ from $P_{F,ij}$, and add $(f_i/0.05)^2$ to χ^2 . This is probably overly conservative for any one bin, but implies a combined freedom $\pm 0.05/3$ (for 9 bins) on an overall noise misestimation. This seems prudent, even though it is not really required by any test we have performed.

2.4. Accuracy of the Resolution

The resolution of the SDSS spectra is estimated using lines from calibration lamps mounted on the telescope structure. Shifts of the detector pixel grid relative to a fixed observed wavelength frame during an exposure, which we will call flexure, are expected to be the dominant source of error in this spectral resolution estimate. We tried estimating the rate of shifting for each pixel by differencing the wavelength calibrations of adjacent exposures (this calibration is determined very precisely for each exposure using the positions of sky lines). The implied extra smoothing only changes the power by $\sim 2\%$ at our highest k bin.

The strong sky line at 5577 Å provides a good opportunity to measure the resolution more directly (note that the spectral wavelengths are in vacuum, and heliocentric, so this and other sky lines generally appear shifted from their standard wavelength). We measure the power spectrum in ~ 3000 sky spectra in the range $5560\text{Å} < \lambda < 5598\text{Å}$. If the sky line has negligible width and the smoothing has a Gaussian shape with rms width R , the power spectrum should be proportional to $W^2(k, R, l) = \exp[-(kR)^2][\sin(k l/2)/(k l/2)]^2$, where l is the pixel width (the pixelization effect is subdominant but not negligible). In Figure 5 we show the measured power averaged over all the sky spectra after dividing each individual measurement by $W^2(k, R, l)$, where R and l are the local values (they are to a good approximation constant over the range we are looking at), and also

dividing each measurement by the value at a low k where the resolution should not have any effect. The result is remarkably close to unity, indicating that the estimated resolution is an accurate representation of the true resolution. What are the small wiggles? Figure 6 shows an example of the region we Fourier transform to measure the power. We believe the small features to the sides of the main line are OH lines at 5564 and 5589 Å (Slanger et al. 2003). We test this explanation for the wiggles by constructing mock sky spectra that simply have a delta function at 5579 Å and two more with 0.003 times the main line’s amplitude at 5566 and 5591 Å, convolved with the resolution and pixelization (0.003 was chosen to give the best fit to the wiggles). The red, dotted line in Figure 5 shows the same resolution test using the mocks. We see that the wiggles are essentially perfectly reproduced. In conclusion: the resolution profile appears to be perfectly Gaussian, with exactly the width expected from the given resolution. There is apparently no room for even a 2% level effect from flexure. We are prevented from performing the same kind of measurement using other sky lines by similar features which are always much larger relative to the central line.

We suggest that fits to $P_F(k, z)$ include a multiplicative uncertainty on the overall power, of the form $\exp(\alpha k^2)$, where α is a single parameter in the fit subject to the rms constraint $\sigma_\alpha = (7 \text{ km s}^{-1})^2$. This allows for a $\sim 2\%$ change in the smoothing kernel at our highest k , similar to our estimate of the error from flexure. This error estimate is somewhat arbitrary, but the evidence we have presented suggests that it should be smaller, so our estimate is conservative.

Note that this resolution test, and the noise calibration, cannot be used directly with the standard pipeline spectra, where the exposures are combined in a different way. The reader may be confused at this point about how our spectra differ from the standard publicly available set, so we give the following summary:

- Our nearest-grid-point combination of the exposures produces uncorrelated noise in pixels (to the extent that the noise in the exposures was uncorrelated, which is expected from the way they are extracted), while the standard pipeline uses a local splining procedure which does a good but not perfect job of preventing noise correlation.
- When combining exposures we record the effect of different pixel sizes, misalignment of the pixels, and flexure of the detector during exposures, which can influence the effective resolution.
- We record the contribution of quasar flux, sky flux, and read-noise to the total noise in each pixel. Knowing the contribution from quasar flux is important if pixel-by-pixel noise weighting is to be used, because the correlation between flux level and noise amplitude can lead to biases (see the end of §2.5).
- We correct for the bias in the exposure combination associated with cross-correlation between the noise variance level in exposure pixels and the quasar flux in the pixel (a different incarnation of the problem alluded to in the previous point).

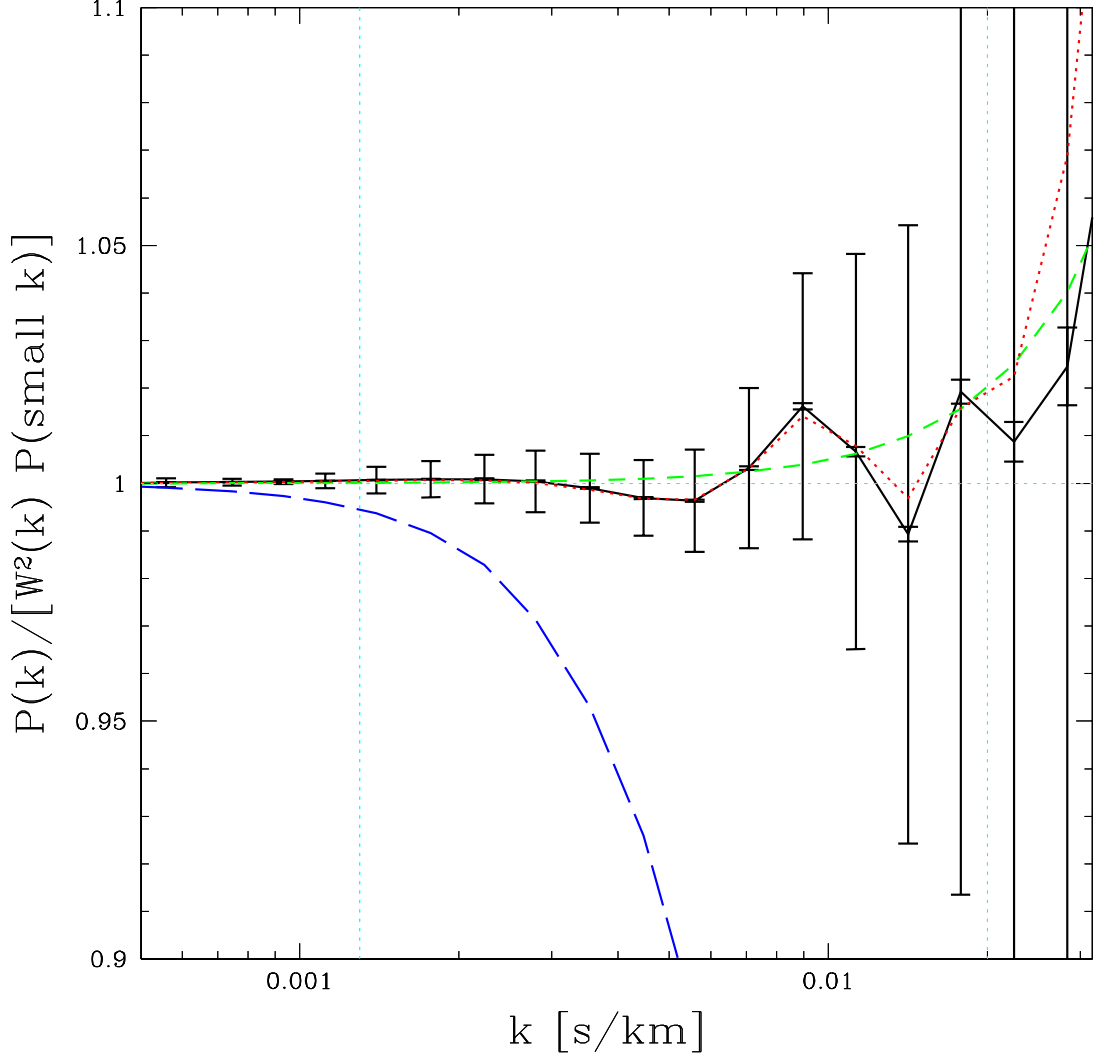


Fig. 5.— Resolution test. The solid, black line with error bars shows the power measured in ~ 3000 sky spectra in the range $5560\text{\AA} < \lambda < 5598\text{\AA}$ (dominated by the strong sky line at 5577\AA) divided by the asymptotic small k power and by the estimated resolution/pixelization kernel $W^2(k)$ for each spectrum. If the resolution estimate was perfect, and the sky line was narrow and the only flux present, this division would give exactly 1. The large error bars are the spectrum-to-spectrum variation, the small ones are the error on the mean. The blue, long-dashed line shows the power not divided by $W^2(k)$, i.e., basically an averaged version of $W^2(k)$, which drops to ~ 0.25 by $k = 0.02(\text{km s}^{-1})^{-1}$. The red, dotted line shows the result of our test for mock spectra constructed with a Gaussian at 5579\AA and two more at 5566 and 5591\AA with 0.003 times its amplitude, representing OH lines. The green, short-dashed line shows $\exp[(k/7\text{ km s}^{-1})^2]$. The vertical, cyan, dotted lines bound the k region in which we will present Ly α forest results, while the horizontal, cyan, dotted line just guides the eye to 1.

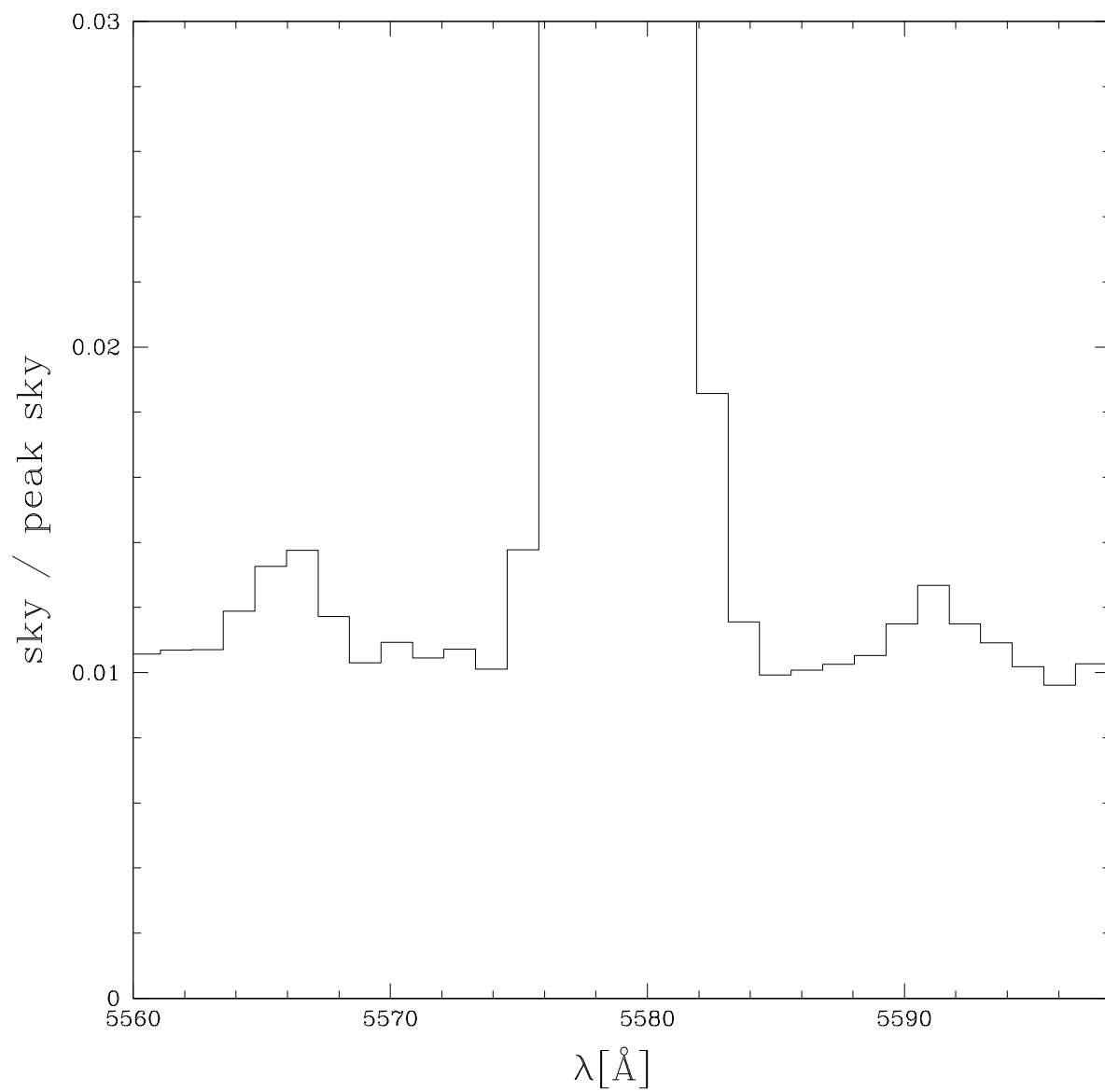


Fig. 6.— Example of the sky flux near the sky line at vacuum wavelength 5579Å, relative to the peak of the line (one spectrum only).

- The noise is recalibrated for each spectrum by differencing the exposures. The noise variance in the standard pipeline exposure spectra is underestimated by on average 16%, on top of any error related to the exposure combination, and the power measured in the difference spectra is slightly tilted relative to white noise.

The last point is the most important.

2.5. Determination of the Continuum and Mean Absorption Level

Our goal is to measure the power spectrum of the fluctuations in the transmitted flux fraction through the IGM, $\delta_F(\lambda) = F(\lambda)/\bar{F} - 1$, where $F(\lambda) = \exp[-\tau(\lambda)]$ and $\tau(\lambda)$ is the Ly α forest optical depth (as defined in §1). However, the spectrum of each quasar is the product of F and the quasar continuum (note that we use “continuum” to refer to all the flux emitted by the quasar, including emission lines), further complicated by errors in the detector calibration and absorption by longer wavelength transitions. The details of the procedure we use to separate these contributions will be presented elsewhere, here we give the basic idea and key results that are relevant for the flux power spectrum determination.

We use an iterative procedure to determine the components of the data model

$$f^i = A_q \bar{C}(\lambda_{\text{rest}}^i) (1 + \delta_C^i) \bar{S}(\lambda^i) (1 + \delta_S^i) \bar{F}(z^i) (1 + \delta_F^i) + n^i, \quad (2)$$

where f^i is the raw flux in pixel i , n^i is the noise, A_q is the overall normalization of the quasar spectrum, $\bar{C}(\lambda_{\text{rest}})$ is the mean quasar continuum shape, δ_C are fluctuations around the mean continuum, $\bar{S}(\lambda)$ is a mean generalized calibration vector (this includes wavelength dependent calibration errors in the SDSS spectra, but also the mean absorption by metal lines with resonance wavelength $\lambda \gtrsim 1300\text{\AA}$), δ_S are fluctuations around \bar{S} , such as individual metal lines or variable calibration errors, $\bar{F}(z)$ is the mean Ly α forest absorption at a given redshift, and δ_F are the fluctuations in Ly α forest absorption. Note that here, as most places in this paper, $z^i = \lambda^i/\lambda_\alpha - 1$ is the redshift of gas that would produce Ly α absorption in the pixel, not the redshift of the quasar. Briefly, we determine A_q for each spectrum, the global functions $\bar{C}(\lambda_{\text{rest}})$, $\bar{S}(\lambda)$, and $\bar{F}(z)$, and a set of principal component analysis (PCA) eigenvectors that describe δ_C by, in turn, computing each component of the model from all the spectra while holding all the others fixed. E.g., $\bar{F}(z)$ is estimated in bins of z by averaging $f^i/A_q \bar{C}(\lambda_{\text{rest}}^i) \bar{S}(\lambda^i) (1 + \delta_C^i)$ over all the pixels in the Ly α forest that fall in each bin. We separate $\bar{S}(\lambda)$ from $\bar{F}(z)$ by measuring $\bar{S}(\lambda)$ in the rest wavelength range $1268\text{\AA} < \lambda_{\text{rest}} < 1380\text{\AA}$, i.e., outside the Ly α forest. A few iterations suffice to determine all the components of the model independently. Three examples of the results are shown in Figure 3. The full details of this procedure will be presented in a separate paper focused on a precise determination of $\bar{F}(z)$.

In preparation for measuring the power spectrum, we divide each quasar spectrum by our estimate of $A_q \bar{C}(\lambda_{\text{rest}}) (1 + \delta_C) \bar{S}(\lambda) \bar{F}(z)$. The power in the $\bar{S}(\lambda)$ and $\bar{F}(z)$ terms is completely

negligible (always $< 0.5\%$ of $P_F(k, z)$ and usually much less). δ_C are represented for each quasar by N PCA eigenvectors. We have tried several different values for N ranging from 0 to 13, and find that the power spectrum results depend slightly (but not critically) on the value, as we discuss below. For our final results we use $N = 0$, i.e., we only divide by a mean continuum, although we will also show how using $N = 13$ affects the cosmological fit results. We do not use the PCA continua because we are not satisfied with their robustness, and division by them frequently actually increases the resulting power slightly. This may indicate that the error introduced by allowing additional freedom in the continuum is larger than the continuum fluctuations that we are trying to remove. Our study of PCA continua was primarily aimed at determining $\bar{F}(z)$ rather than the power spectrum, so we cannot be certain that the PCA method could not be used productively in a power spectrum measurement if it was more carefully optimized for that application. Because we know that our continuum estimate (which involves an extrapolation from outside the Ly α forest region) is not perfect within the Ly α forest, we further divide each chunk of spectrum that will be used to make a power spectrum estimate by its mean (optimally computed considering both observational noise and absorption variance). We call our resulting observed data vector $\delta_f = \delta_F + \delta_S + \delta_n$, where δ_n is the normalized noise fluctuation and we are ignoring the cross-terms between δ_F , δ_C , and δ_S . As we describe in detail below, δ_S is treated as a random noise background and its statistical properties are determined by measuring the power spectrum in the rest wavelength range $1268 < \lambda_{\text{rest}} < 1380\text{\AA}$, where $\delta_F \equiv 0$ (and $\bar{F} \equiv 1$).

A small but non-negligible detail of our procedure is hidden within our description of the normalization of the spectra. When we estimate the mean to divide by, we weight the computation optimally using the covariance matrix, C_{ij} , of the pixels (C is discussed in more detail in our explanation of the power spectrum measurement below). C includes Ly α forest fluctuations and measurement noise. We do not use our best estimate of the measurement noise directly for the weighting, because the noise variance estimate is correlated with the measured flux in the pixel, which leads to a bias: the mean is underestimated because lower flux pixels have lower noise. The original noise estimate is $\sigma_p^2 = \gamma (f_{\text{quasar}} + f_{\text{sky}}) + \sigma_{\text{readnoise}}^2$, where f_{quasar} is the flux from the quasar, f_{sky} is the flux from the sky, and γ accounts for the conversion between the units of flux and photons (this description is slightly idealized since the reduction of two-dimensional CCD data to a spectrum introduces some complications). To remove the correlation between flux and noise, we subtract γf_{quasar} from σ_p^2 and add $\gamma \langle f_{\text{quasar}} \rangle$, where $\langle f_{\text{quasar}} \rangle = A_q \bar{C}(\lambda_{\text{rest}}) (1 + \delta_C) \bar{S}(\lambda) \bar{F}(z)$. We call the final result σ_w (w for weight; see Figure 3 for some comparisons of the noise estimates). The estimate of γ we have from the spectral reduction pipeline is not perfect, so our replacement of the correlated part of the noise amplitude is imperfect. We make a final, very small, correction to the mean estimation based on a direct computation of the cross-correlation between the flux and noise amplitude. We use the same decorrelated noise amplitudes for weighting the power spectrum extraction (discussed below); however, the bias is completely insignificant in that case (i.e., $P_F(k, z)$ computed using the original noise estimates for weighting is practically identical).

3. Power Spectrum Determination

The high precision of the $P_F(k, z)$ measurement obtainable using the SDSS data sample requires unprecedented (in this field) care in the design and testing of the procedure used to produce it. We describe the basic method that we use to extract a power spectrum and estimate the errors in §3.1. In §3.2 we present the test of the full method as implemented in our code, using mock data sets. In §3.3 we give the raw result for the measurement of power in the Ly α forest region.

We aim to measure the power spectrum of δ_F , representing the correlation of fluctuations in the Ly α forest absorption only; however, the covariance matrix of the data vector δ_f is

$$\langle \delta_f \delta_f^T \rangle = \langle \delta_F \delta_F^T \rangle + \langle \delta_S \delta_S^T \rangle + \langle \delta_n \delta_n^T \rangle \quad (3)$$

(the three components of δ_f as we have defined it should be uncorrelated). The noise term in equation (3) is relatively easy to compute and subtract. We estimate and subtract most of $\langle \delta_S \delta_S^T \rangle$ by defining

$$P_F(k, z) = P_{1041,1185}(k, z) - P_{1268,1380}(k, z) , \quad (4)$$

where z is always defined by $z = \lambda/\lambda_\alpha - 1$, so that we are subtracting power measured in the same observed wavelength ranges, not the same quasar spectrum (we remind the reader that we have defined P_{λ_1, λ_2} to mean the power measured in the range $\lambda_1 < \lambda_{\text{rest}} < \lambda_2$). As it appears in $P_{1268,1380}(k, z)$, z is the redshift of gas that would produce Ly α absorption in this part of the quasar spectrum, if it was not at a higher redshift than the quasar, i.e., z is really just an indicator of observed wavelength. The subtraction in equation (4) will completely remove the power due to transitions with $\lambda > 1380$ Å, including SiIV (a doublet absorbing at rest wavelengths 1393.75 and 1402.77 Å) and CIV (another doublet at 1548.20 and 1550.78 Å). Note that this subtraction of metal power is exact, not an approximation [except for the approximation that $(1 + \delta_F)(1 + \delta_S) \simeq 1 + \delta_F + \delta_S$], because we are determining the metal power in exactly the same observed wavelength range as the Ly α forest power from which it is being subtracted, i.e., the same gas, at the same redshift, is doing the absorbing both inside and outside the forest, so the absorption will have identical statistical properties. This background subtraction will also remove any strictly observed-wavelength-dependent power introduced by the detector, such as spectrum to spectrum variations in the calibration of the detector. We implement it in §3.4.

3.1. Core Method

In this subsection we describe our method for extracting the power spectrum, $P_X(k, z)$, from any selected rest wavelength range X (§3.1.1), and estimating its statistical uncertainty (§3.1.2).

3.1.1. Band-Power Estimation

We estimate $P_X(k)$ using the quadratic estimation method, which is essentially a fast iterative implementation of the maximum likelihood estimator (we follow the expressions as given in Seljak (1998)). This method is optimal for a Gaussian probability distribution. While the power spectrum estimates are not Gaussian distributed, the deviations are small, as shown below. We measure the power in flat bands with edges given by $\log_{10}(k_i) = -4.2 + 0.1 i$ where i ranges from 0 to 30 (to produce 29 bands), although we will not give results for some of the large and small-scale bands when we think they are unreliable. Defining $\delta_f = \delta_X + \delta_n$, where δ_X are the fluctuations we are measuring (e.g., $\delta_X = \delta_F + \delta_S$ within the forest) and δ_n are the normalized noise fluctuations, a band-power estimate, \hat{P}_k , for each chunk of spectrum is given by

$$\hat{P}_k = \frac{1}{2} \sum_{k'} F_{kk'}^{-1} (\delta_f^T C^{-1} Q_{k'} C^{-1} \delta_f - b_{k'}) , \quad (5)$$

where $C = \langle \delta_f \delta_f^T \rangle = S + N$, $S = \langle \delta_X \delta_X^T \rangle$, $N = \langle \delta_n \delta_n^T \rangle$, $Q_k = \partial S / \partial P_k$,

$$F_{kk'} = \frac{1}{2} \text{tr}(C^{-1} Q_k C^{-1} Q_{k'}) \quad (6)$$

is the Fisher matrix and the noise bias is

$$b_k = \text{tr}(C^{-1} Q_k C^{-1} N) . \quad (7)$$

Note that we could include the background power explicitly in these equations as a noise source when measuring the power in the Ly α forest region, but we ignore this because its contribution is too small to change the weighting significantly. We will subtract it from the estimates later. The noise subtraction term, b_k , is computed using the pipeline noise estimates, σ_p (not σ_w), with the amplitude corrected as discussed above based on the differences between exposures. In principle, S in these equations should be the true covariance matrix; however, as we discuss below, we use the measured covariance from a previous iteration of the power spectrum determination instead.

Except in a few cases that we will identify as they arise, when we set out to measure the power in a defined rest-wavelength region (e.g., $1041 < \lambda_{\text{rest}} < 1185 \text{ \AA}$ for the Ly α forest region) we first use equation (5) to estimate the power separately in halves of the region in each spectrum (e.g., $1041 < \lambda_{\text{rest}} < 1113 \text{ \AA}$ and $1113 < \lambda_{\text{rest}} < 1185 \text{ \AA}$). Our choice of half-spectra is a compromise between competing desires for resolution in redshift and wavenumber. The full length of the forest in a spectrum corresponds to a redshift interval, $\Delta z \simeq 0.4$, that is unnecessarily large. While the precision of the measured power spectrum would support smaller than half-spectrum chunks to give finer redshift resolution than $\Delta z \simeq 0.2$, the shorter chunks would limit the k -space resolution. Note that we could have used full chunks and still achieved the same z -resolution by more carefully applying the estimator equation, as we discuss below, but this would increase the computational time without much improvement in the final errors on the scales of relevance.

After computing estimates \hat{P}_k for each half-spectrum, we perform a weighted average to determine $P_X(k, z)$ in redshift bins centered at $z_i = 2 + 0.2 i$ where $i = 1..13$ (in this paper we only

present results up to $i = 9$). Each bin is the average of the power in all the half-spectra for which the redshift of the central pixel falls within ± 0.1 of the bin center (we discuss below how we correct for an asymmetric distribution of data within a bin). We combine sets of estimates using the Fisher matrices (equation 6) for the weighting. In practice this means that we sum the quantity in parentheses in equation (5) over all estimates and multiply the result by the inverse of the sum of the Fisher matrices for each individual estimate. Our procedure would be optimal for Gaussian data, which the Ly α forest is not; however, when we use the Gaussian approximation to compute the errors on the measured power the results are not much different from the more accurate bootstrap errors (see §3.1.2), so we conclude that our method is not far from optimal.

Whenever we have a finite length of spectrum, there will be mixing between the power in different bins. Variable noise or gaps in the data will produce more mixing. This mixing is described in terms of a window matrix, which is given by the Fisher matrix in equation 6. In our standard procedure, the power spectrum estimates in equation 5 are multiplied with the inverse of Fisher matrix and are thus deconvolved with the window, which removes the mixing of other modes into the bin one is estimating (however, the bins are still correlated). This method thus produces a diagonal window matrix, so each combined estimate of $P_X(k, z)$ represents exactly the range of k corresponding to its bin. Our tests below show that there is no practical problem with instability in the Fisher matrix inversion (the window matrices are close to diagonal to begin with). A diagonal window matrix is desirable from a theoretical standpoint because our ability to compute the power spectrum from simulations is limited at both low k (by limited box size) and high k (by simulation resolution and complexity of physics). In the few cases where we use the power without deconvolution, we are using the estimator $\mathcal{N}_k (\mathbf{F} \mathbf{P}_X)_k$, where $\mathcal{N}_k \equiv (\sum_{k'} F_{kk'})^{-1}$ (Seljak 1998).

To compute the weight matrix C , we need an estimate of S , i.e., the power spectrum we are trying to measure. We solve this problem by computing $P_X(k, z)$ iteratively. The first estimate is made assuming $S = 0$. In subsequent iterations we compute S from the previous estimate of $P_X(k, z)$. This procedure converges quickly (the difference between $S = 0$ and a reasonable estimate of S is significant, but once S is in the right ballpark it does not matter what it is exactly). We add a large constant (10.0) to all elements of the weight matrix, to remove all direct sensitivity of our power measurement to the mean of the chunk. This makes very little difference to the results on the scales we present. We are however still sensitive to the mean estimate from when we divided the spectrum by it. Even if the mean estimates are correct on average, the statistical error on the mean for each spectrum can still lead to a bias. If the errors on the mean estimate were small and uncorrelated with the fluctuations in the flux field, the bias would be $1 + 3 \sigma_m^2$, where σ_m is the error on the mean [to lowest order in σ_m , i.e., the bias is $\langle 1/(1 + \delta_m)^2 \rangle \simeq 1 + 3 \langle \delta_m^2 \rangle$, where δ_m is the fractional error in the mean, and $\langle \delta_m^2 \rangle = \sigma_m^2$]. We divide each estimate by this factor as part of our standard procedure; however, as we discuss below when we test our code on mock spectra, this approximation is not sufficient and we will need to include another small, k -dependent, factor determined numerically using the mock spectra (this is the only use of the mock spectra other than for testing).

The reader may at this point be wondering what redshift the resulting $P_X(k, z_i)$ should be taken to represent, i.e., z_i is not necessarily the center of weight of the data, and neither is the mean redshift of the pixels in the bin, considering the rather complicated weighting in equation (5). In fact, the effective redshift is not even the same for each k -bin in the same z -bin. We resolve this question – $P_X(k, z_i)$ represents the power spectrum at precisely z_i (to first order) – in our construction of $S^{ab} = \langle \delta_X^a \delta_X^b \rangle$ and $Q_k^{ab} = \partial S^{ab} / \partial P_{k,i}$, where a and b label pixels at redshifts z_a and z_b , and i labels the redshift bin in which this chunk of spectrum falls. To account for the evolution from z_a and z_b to z_i , we define a power spectrum growth factor, $D_{k,i}(z) = [(1+z)/(1+z_i)]^{\alpha_{k,i}}$, where

$$\alpha_{k,i} = \frac{d \ln[P_X(k, z)]}{d \ln(1+z)} \Big|_{z_i} \simeq \frac{\ln(P_{k,i+1}/P_{k,i-1})}{\ln[(1+z_{i+1})/(1+z_{i-1})]} \quad (8)$$

(we use a one-sided derivative estimate instead of equation 8 for the first and last redshift bins). Now $Q_k^{ab} = D_{k,i}(z_{ab}) Q_k^{ab}|_{z_i}$, where $z_{ab} = (z_a + z_b)/2$ and $Q_k^{ab}|_{z_i}$ is computed as if the pixels were located at the center of the bin. Finally, $S^{ab} = \sum_k Q_k^{ab} P_k(z_i)$. This correction may be difficult to understand intuitively at first, but it is really quite simple. The modification of Q just corrects the power spectrum estimate for the excess (dearth) of power that we expect for pixels in the high (low) redshift ends of the bin. The correction to S affects the weighting, simply producing a more accurate S at the redshift of the pixels in question.

Note that an alternative method would be to treat the points $P_X(k, z_i)$ as simply parameters of a continuous power spectrum defined by some form of interpolation. This would mean S^{ab} would have non-zero derivative with respect to more than one of the power spectrum bins (e.g., usually two for linear interpolation). This method would be elegant, and probably produce narrower effective window functions in the z direction; however, it will increase the correlation in the z direction between measurement errors, because the same pixels would contribute to more than one power spectrum point. Since this more sophisticated method would allow long chunks of spectra to be used without degrading our z resolution, it would be most useful if we were trying to measure the power on even larger scales.

How does our method compare to the straightforward Fourier transform (FT) method? The basic FT method is to project the data vector, δ_f , onto a set of modes of the form $d_{\alpha\beta} = \exp(i k_\alpha \Delta v_\beta)$, and to simply compute the variance of the amplitudes of all the modes with k in some bin, i.e.,

$$\hat{P}_{\text{FT}} \propto \sum_{\alpha}^{k_{\min} < |k_\alpha| < k_{\max}} \left| \sum_{\beta} d_{\alpha\beta} \delta_{f,\beta} \right|^2, \quad (9)$$

where k_{\min} and k_{\max} define the bin, and the discrete spacing of k_α is somewhat arbitrary (the natural spacing is $\Delta k = 2\pi/L$, where L is the length of the spectrum, but nothing prevents one from choosing more finely spaced k s). Our estimator, equation (5), can be cast in a similar form, i.e., as a projection of the data vector onto a set of modes, and a sum of the squares of the mode amplitudes. We require that the mode amplitudes are statistically independent, which makes their computation equivalent to a computation of Karhunen-Loève eigenmodes (see, e.g., Tegmark et al.

(1997)). Figure 7 shows the two most important modes for our bin with $0.00126 \text{ (km s}^{-1}\text{)}^{-1} < k < 0.00158 \text{ (km s}^{-1}\text{)}^{-1}$, for the chunk of spectrum shown in Figure 3a. In this case two modes differing primarily by a phase shift, analogous to $\sin(kx)$ and $\cos(kx)$, contain most of the information, because our bin width is approximately $2\pi/L$. We see that the difference between our modes and a simple sine wave is not dramatic – there is a little bit of edge tapering (downweighting the edges to make the effective window on the data more compact in Fourier space) and some straightforward downweighting of the most noisy pixels. Curiously, there seems to be an additional effect where pixels adjacent to an edge are given extra weight, possibly as a way of compensating for missing data (this is seen more clearly in spectra where a narrow gap is present in the middle of the data). The picture is similar for bins with larger k , except of course that there are increasingly many important modes as the width of our bins increases (the bins have a fixed width $\Delta \log(k)$, but the relevant mode width is Δk). For more discussion of the quadratic estimator see, e.g., Tegmark (1997).

The method we adopt is optimal for Gaussian fields and therefore guarantees that no other method can surpass it. An additional advantage is that within this formalism window and covariance matrices are automatically computed. For continuous spectra with few gaps and near uniform noise one does not necessarily expect an FT method to be significantly worse. In practice the noise level is slowly varying across the spectrum, so averaging all the pixels uniformly is not optimal and degrades performance. Another advantage is that with our method each pixel pair has its own effective redshift and the correlations for a given pair are then interpolated to the redshift of interest using the appropriate evolution. In the FT method the whole spectrum is Fourier transformed first, so the redshift information is preserved only in an averaged sense, but a priori it is not clear how this average is defined.

3.1.2. Bootstrap Error Estimation

While the Fisher matrix obtained during the estimation process would give the error matrix for $P_X(k, z)$ if the data were Gaussian, we cannot reliably assume this. Our solution is to compute a bootstrap error matrix by the standard procedure (Press et al. 1992). From our data set of N spectra, we form a bootstrap data set by selecting N spectra at random, with replacement. The covariance matrix of $P_X(k, z)$ is taken to be $M^{ij} = \langle \Delta \tilde{P}_X^i \Delta \tilde{P}_X^j \rangle$, where $\Delta \tilde{P}_X^i = \tilde{P}_X^i - \langle \tilde{P}_X^i \rangle$, \tilde{P}_X^i is an estimate of the power in the i th bin from a bootstrap data set, and $\langle \rangle$ means average over bootstrap realizations. We generally use 4000 realizations, after checking that this produces convergence in the result. We assume that the error correlations extend only one bin off diagonal in the z direction, because the spectrum of a single quasar practically never contributes to non-adjacent bins.

We have no compelling reason to believe that this method of computing the error bars will give rigorously correct results. Considering the large number of off-diagonal elements that must be

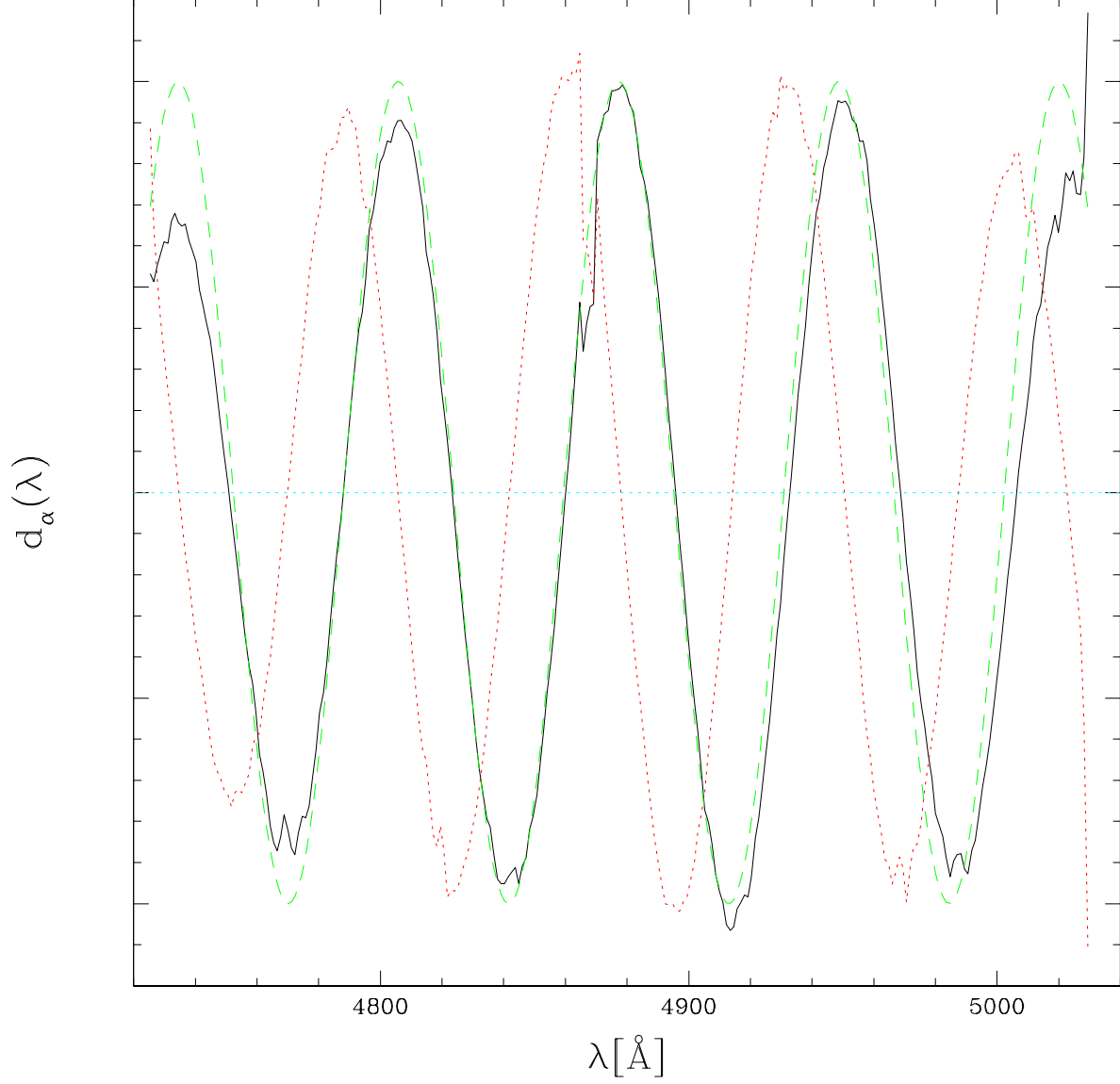


Fig. 7.— Black solid and red dotted lines show the two primary modes onto which the data is effectively projected when we estimate the power in our bin centered on $k = 0.00141 \text{ (km s}^{-1}\text{)}^{-1}$, for the spectrum shown in Figure 3a. The horizontal axis scale in the figure is arbitrary. For comparison, the dashed line shows a simple sine wave with $k = 0.00141 \text{ (km s}^{-1}\text{)}^{-1}$.

estimated, one worry is that a particular linear combination of the bins may accidentally vary very little in our data set, so it will appear to have an unrealistically small error. Our tests on mock spectra (§3.2.2) show no sign of this problem. Still, to be conservative we apply one tweak to M after it is computed, in an attempt to inoculate it against the possibility. We perform a singular-value decomposition on M , which produces a set of independent vectors and their variances. We then compute the variances of the same vectors under the Gaussian approximation, using the Fisher matrix. If the bootstrap variance is smaller than the Gaussian variance we replace it with the Gaussian variance. Finally, we transform back to M . The tests on mock samples described below give us confidence that our procedure is reliable.

3.2. Tests on Mock Data Sets

We validate our procedure as implemented in code by applying it to mock data sets. Many iterations of these tests were required to produce results that show no serious problems in the error estimation or the power spectrum estimation itself. Testing the results on realistically created mock data is absolutely essential for measurements of such high precision. In §3.2.1 we describe our procedure for generating the mock spectra. We test our bootstrap error estimates in §3.2.2. Finally, we test the power spectrum estimation procedure for systematic errors in §3.2.3.

3.2.1. Generating Mock Spectra

We generate mock spectra by combining the auxiliary information we have for each observed spectrum (e.g., our continuum estimate, noise estimate, sky estimate, etc.) with a simplified version of the Bi et al. (1992) model for the $\exp(-\tau)$ field, which results in realistic looking spectra.

For each observed quasar we start with the term we divide by before computing the power spectrum, $A_q \bar{C}(\lambda_{\text{rest}}) (1 + \delta_C) \bar{S}(\lambda)$ (see equation 2). We multiply this by $\exp(-\tau)$ (generated as described below), smooth the result using the resolution from the observed spectrum, and sample the result onto the observed grid of pixels. This produces a noise free version of the flux we would observe coming from this quasar. We add flux from the sky as estimated for the observed spectrum, and transform the total flux to the number of photons that would be expected in each pixel. We generate a Poisson deviate with this mean, add the appropriate Gaussian read-noise for each pixel, transform back to the original flux units, and subtract the sky flux estimate to obtain an observed (noisy) quasar spectrum. The results of this procedure for each observed quasar are written into files in the same format as the observed spectra, so exactly the same code can be used to measure the power in the mock spectra.

To generate the $\exp(-\tau)$ fields we use a simple model that is arranged to give roughly the correct power spectrum as a function of k and z , and the correct mean absorption as a function of redshift. For each observed spectrum, we start by generating a Gaussian random field, $\delta_{i,0}$, on a

very long, relatively finely spaced grid (65536 cells with width 7 km s^{-1} , to be precise), with power spectrum

$$P_\delta(k) = \frac{1 + [0.01 (\text{km s}^{-1})^{-1}/k_0]^\nu}{1 + (k/k_0)^\nu} \exp[-(kR_\delta)^2], \quad (10)$$

where $k_0 = 0.001 (\text{km s}^{-1})^{-1}$, $\nu = 0.7$, and $R_\delta = 5 \text{ km s}^{-1}$ [this P_δ was chosen after some experimentation because it produces a final flux power spectrum with approximately the same k dependence as the observed $P_F(k, z)$]. An arbitrary cell in this grid is chosen to correspond to the redshift of the quasar, and the evolution of the amplitude of the power spectrum with redshift is imposed by the transformation $\delta_i = a(z_i) \delta_{i,0}$ with $a^2(z_i) = 58.6 [(1 + z_i)/4]^{-2.82}$, where the form of $a(z)$ was chosen so that the final flux power spectrum would evolve like the observed one. Next we make the squared lognormal transformation $n_i = [\exp(\delta_i - \sigma_i^2/2)]^2$, where σ_i^2 is computed from the input power spectrum, including the amplitude factor (the factor $\sigma_i^2/2$ in the exponential just fixes the mean of the lognormal field to 1). We smooth the n field with a Gaussian filter with rms width $R_\tau = 20 \text{ km s}^{-1}$ and multiply it by a factor $0.374 [(1 + z_i)/4]^{5.10}$ to produce a field τ (this redshift evolution factor produces roughly the observed redshift evolution of \bar{F}). The mock transmitted flux in each grid cell is then $F_i = \exp(-\tau_i)$, which is sampled as described above.

The procedure described above leads to realistic looking spectra of the Ly α forest. We have verified that it generates a bispectrum that is within a factor of 2 of the one measured in N-body simulations. The main advantages of this procedure over the N-body simulation approach when generating the mock spectra are that it is faster, so one can make an arbitrary number of independent realizations, and that the simulated spectra can be of arbitrary length, important to eliminate any periodicity effects (this would be impossible with simulations, where a typical box size is much shorter than the total length of a single spectrum). Both of these advantages are critical for a high precision test. We determine the true $P_F(k, z)$ by a simple FFT of extremely long $\exp(-\tau)$ fields (without redshift evolution).

3.2.2. Tests of the Error Estimates

Without accurate statistical errors it is difficult to identify systematic problems, so we first test our bootstrap procedure for estimating the errors. Note that there is no reason to expect bootstrap errors to be perfect (there is even some ambiguity in how exactly the bootstrapping should be done when the data do not consist of statistically identical objects). Regardless of systematic errors in the method, the only difference between the power spectra measured from two mock data sets that differ only in the random seed that was used to create them should be the statistical errors that we estimate. We test our error bars by generating ten different sets of mock data and computing χ^2 for the differences between each of them and their error weighted mean, using the bootstrap error bars and the 108 points in $0.0013 < k < 0.02 (\text{km s}^{-1})^{-1}$, and $2.1 < z < 3.9$. This is effectively a fit of 108 parameters to 1080 data points, with 972 degrees of freedom. The total χ^2 is 939, perfectly consistent with a random fluctuation around the mean, and strongly disfavoring an underestimation

of the errors by more than a couple percent.

3.2.3. Tests of the Power Estimates

We can now search for systematic errors. To enhance the statistical significance of any errors, we average our ten sets of mock spectra to form a single, more precise measurement. The result is shown in Figure 8. The results look reasonably good; however, we find an unacceptably bad $\chi^2 = 346$ for the comparison between our measured $P_F(k, z)$ and the true power spectrum (there are 108 degrees of freedom). To quantify the systematic problem, we first assume the bias has the form $B(k) = B_0 [k/0.0067 \text{ (km s}^{-1}\text{)}^{-1}]^\nu \equiv P_{\text{measured}}/P_{\text{input}}$ and fit for the values of B_0 and ν that minimize χ^2 in the comparison. We find $B_0 = 1.0036 \pm 0.0014$ and $\nu = -0.0173 \pm 0.0013$ with $\chi^2 = 173$ for 106 degrees of freedom [the pivot point $k_0 = 0.0067 \text{ (km s}^{-1}\text{)}^{-1}$ was chosen to make the errors independent; the amplitude coefficient would be larger if we were not already dividing by $1 + 3 \sigma_m^2$ as explained in §3.1.1]. The combination of slope and amplitude errors corresponds to a 3.1% excess of power at our largest scale, $k = 0.0014 \text{ (km s}^{-1}\text{)}^{-1}$, and a 1.3% underestimate at $k = 0.018 \text{ (km s}^{-1}\text{)}^{-1}$. We find some less significant dependencies by generalizing the fitting formula even more to

$$B(k, z) = B_0 a^\mu(z) \left(\frac{k}{k_0} \right)^{\nu+1/2 \eta \ln(k/k_0) + \zeta \ln[a(z)]}, \quad (11)$$

where $a(z) = (1 + z_0)/(1 + z)$, with $z_0 = 2.85$. The parameters are $B_0 = 1.0073 \pm 0.0016$, $\mu = 0.049 \pm 0.012$, $\nu = -0.0195 \pm 0.0015$, $\eta = -0.0157 \pm 0.0038$, and $\zeta = -0.026 \pm 0.012$, with $\chi^2 = 135$. Where does this bias come from? We expect some bias related to the division of each chunk of spectrum by its overall mean (not because of an integral constraint suppression of large-scale power – our estimator should take care of that – but because of statistical error in the estimate of the mean that we divide by). When we measure the power without this division by the mean, which we can only do using mock spectra, we find significantly smaller corrections – small enough to ignore when model fitting.

We expect that this bias should be present when we use real observed spectra, so we will correct for it by dividing the measured power by $B(k, z)$. We describe its effect on the amplitude and slope of the power spectrum below (table 1).

3.3. Raw Power

Figure 9 shows the raw power measured in our standard Ly α forest rest wavelength range, $1041 < \lambda_{\text{rest}} < 1185$. All the figures in this subsection show $P_{1041,1185}$, not the background subtracted power P_F . Our normalization convention is:

$$\langle \delta^2 \rangle = \int_{-\infty}^{\infty} \frac{dk}{2\pi} P(k). \quad (12)$$

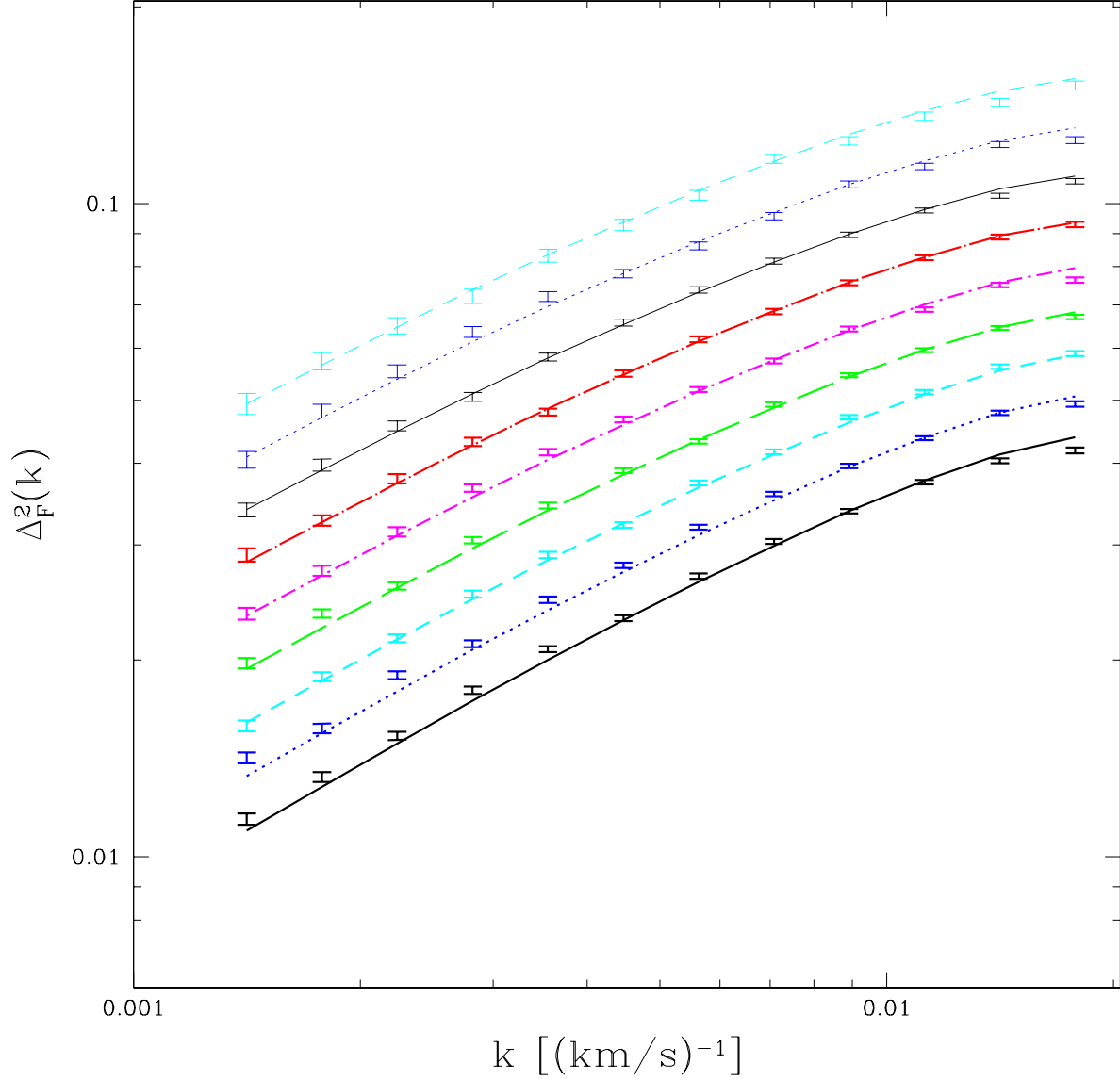


Fig. 8.— Error bands show the average power spectrum [$\Delta^2(k) \equiv \pi^{-1}k P(k)$] measured from ten sets of mock spectra. Lines show the true power. Redshift bins, strictly from bottom to top, are: black, solid line, $z=2.2$, blue, dotted, 2.4, cyan, dashed, 2.6, green, long-dashed, 2.8, magenta, dot-dashed, 3.0, red, dot-long-dashed, 3.2, black, thin, solid line, 3.4, blue, dotted, thin, 3.6, cyan, dashed, thin, $z=3.8$.

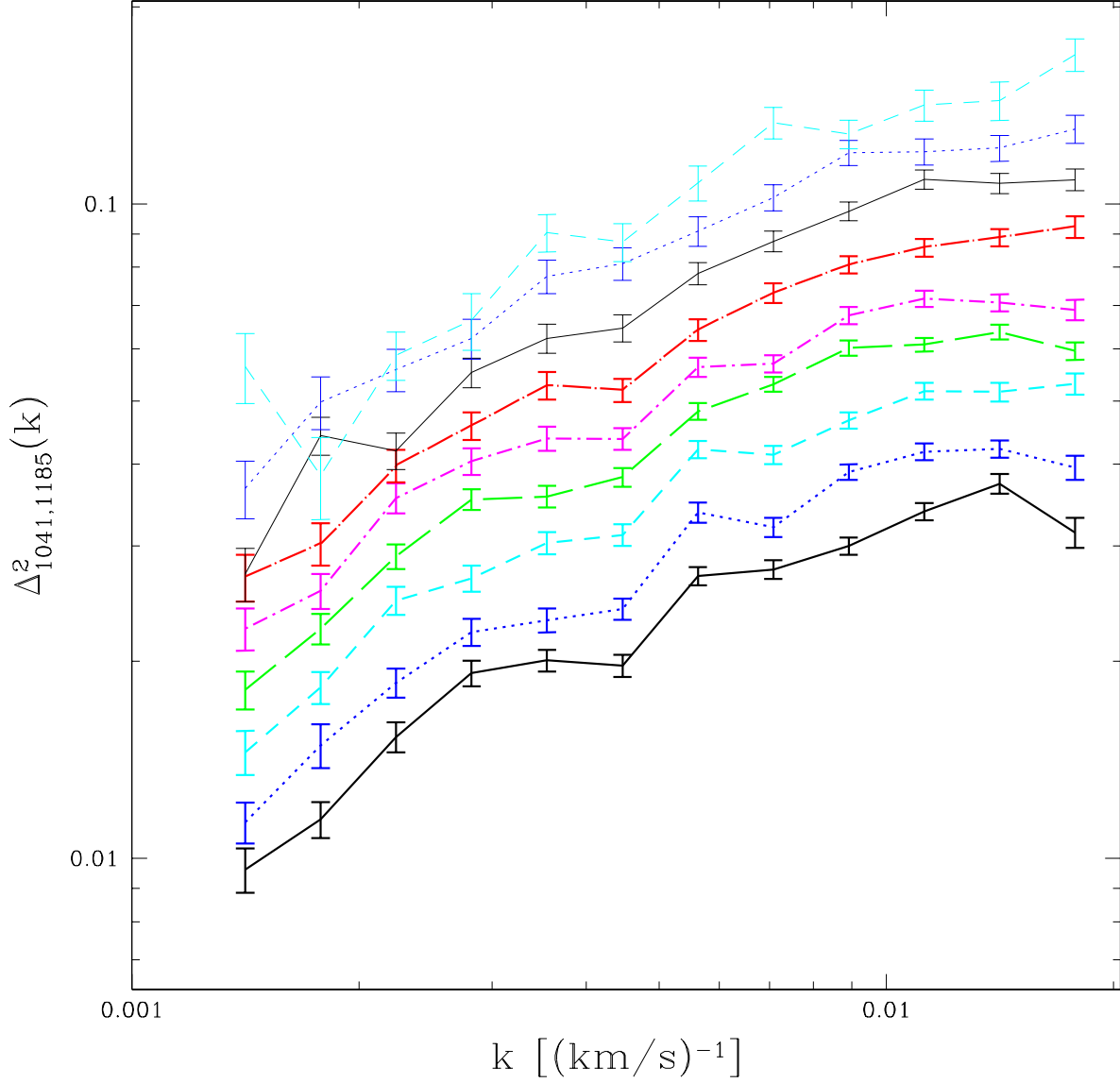


Fig. 9.— Error bars show the power spectrum measured from the observed spectra in the wavelength range $1041 < \lambda_{\text{rest}} < 1185$. The lines connect the points to identify them and to guide the eye. Redshift bins, from bottom to top (roughly) are: black, solid line, $z=2.2$, blue, dotted, 2.4, cyan, dashed, 2.6, green, long-dashed, 2.8, magenta, dot-dashed, 3.0, red, dot-long-dashed, 3.2, black, thin, solid line, 3.4, blue, dotted, thin, 3.6, cyan, dashed, thin, $z=3.8$.

We usually plot the dimensionless quantity $\Delta^2(k) \equiv \pi^{-1}k P(k)$, the contribution to the variance per unit $\ln k$.

Figure 10 shows the fractional errors on all of the measured points. The errors are less than 5% for most of the points, and frequently as small as 3%. If we were only estimating a single amplitude parameter by combining all these points then its error would be 0.6%. An overall logarithmic slope would have an error ± 0.005 . The errors on the largest scales are increased somewhat by the diagonalization of the window matrix.

Figure 11 shows the ratio of subtracted noise power to measured signal power ($P_{1041,1185}$) for each point. The noise power is significant (20-30%) on all scales, but diverges at high k where the resolution suppresses the absorption power. The lowest redshift bin has the most noise, due to the lower Ly α forest power combined with extra noise at the short wavelength end of the spectra.

Figure 12 shows our window matrix (at $z = 2.6$), which we proceed to diagonalize. The matrix is reasonably close to diagonal already, with large contributions only from adjacent bins. It is useful to diagonalize the matrix at this stage, rather than waiting until the model-fitting stage, because this allows us to compute bootstrap errors directly for the final bins (the bootstrap error calculation and window matrix diagonalization do not perfectly commute).

Figure 13 shows the ratio of the bootstrap errors to the errors estimated assuming the data are Gaussian. We did not apply the Gaussian floor to the bootstrap errors when making this figure. Typically the bootstrap errors are 0-20% larger than the Gaussian errors. Figure 14 shows examples of the estimated correlation between the errors, at $z = 2.6$. We note that diagonalizing the window matrix noticeably reduces the error correlations. The bootstrap errors are, in contrast to the Gaussian errors, noticeably correlated ($\langle \Delta P_i \Delta P_j \rangle / \sigma_{P,i} \sigma_{P,j} \sim 0.0-0.2$ when $|i-j| > 1$, where i and j label the bins) across the full k range. These differences between bootstrap and Gaussian errors are not necessarily an indication of intrinsic non-Gaussianity in the absorption fluctuations. Possible alternative explanations for the differences include the uncertainty in the mean flux value that each chunk of spectrum is divided by and the uncertainty in the noise-subtraction term for each chunk, neither of which are included in the Gaussian estimate and both of which would increase the error in a way that is correlated across k bins.

3.4. Background Subtraction

Our background subtraction is the power in the wavelength range $1268 < \lambda_{\text{rest}} < 1380$ Å. Figure 15 shows $P_{1268,1380}$ and $P_{1041,1185}$ for comparison. The bump at $k \sim 0.013$ (km s^{-1}) $^{-1}$ in $P_{1268,1380}$ is probably due to the CIV doublet at separation 499 km s^{-1} . The bump at $k \sim 0.003$ (km s^{-1}) $^{-1}$ may be due to the SiIV doublet at separation 1933 km s^{-1} . Figure 16 shows $P_{1268,1380}/P_{1041,1185}$. We see that, even though the background power is a small fraction of the Ly α forest power, it is quite significant when compared to the small size of the errors on the Ly α forest power. It is important to remember that even a small overall systematic error can be very significant if it covers

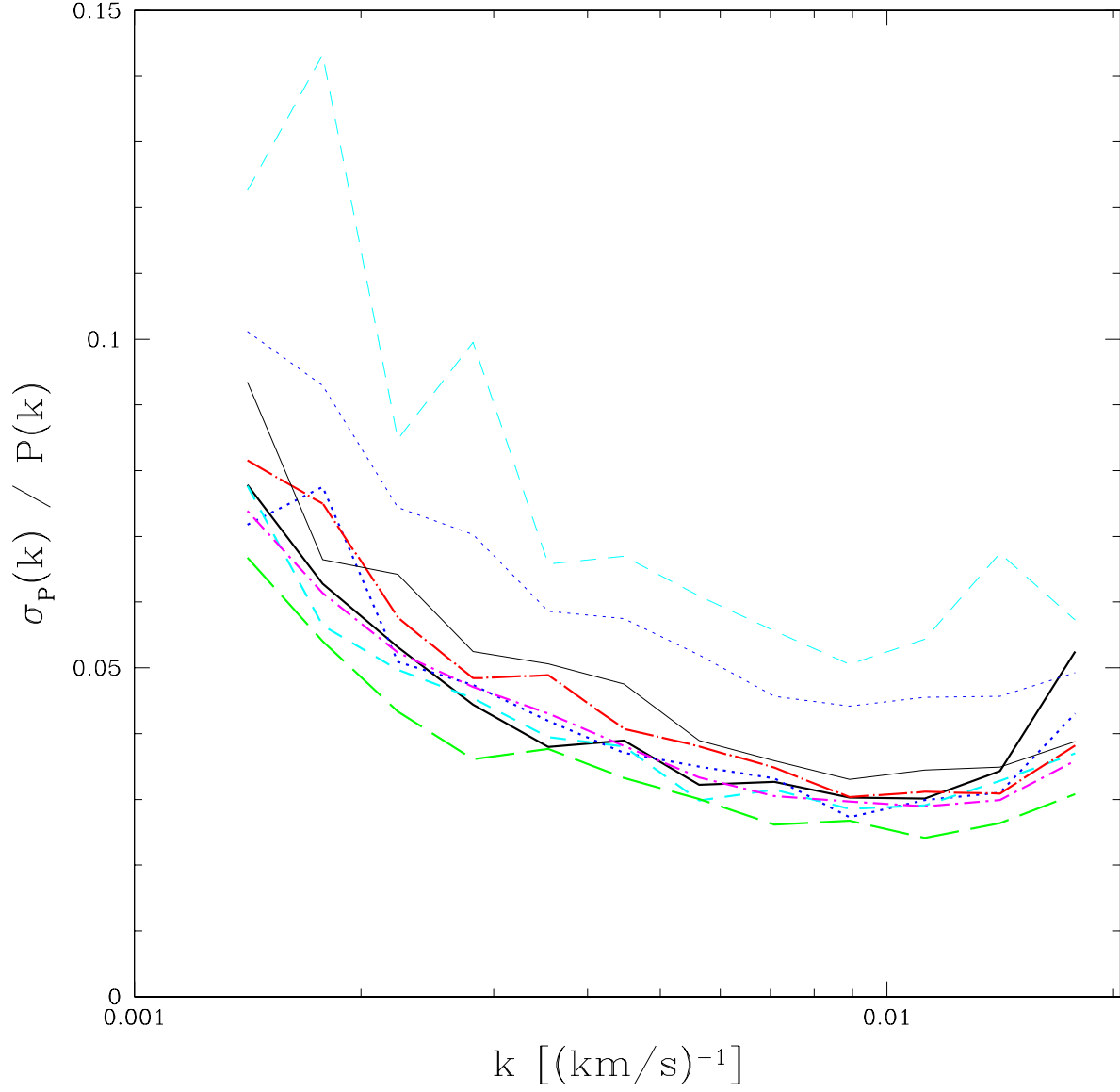


Fig. 10.— Lines connect the fractional errors on each measured $P_{1041,1185}(k, z)$ point, using the usual line-type and color and scheme (see Fig. 9 – the highest two curves are the highest two redshifts, the lowest is $z = 2.8$).

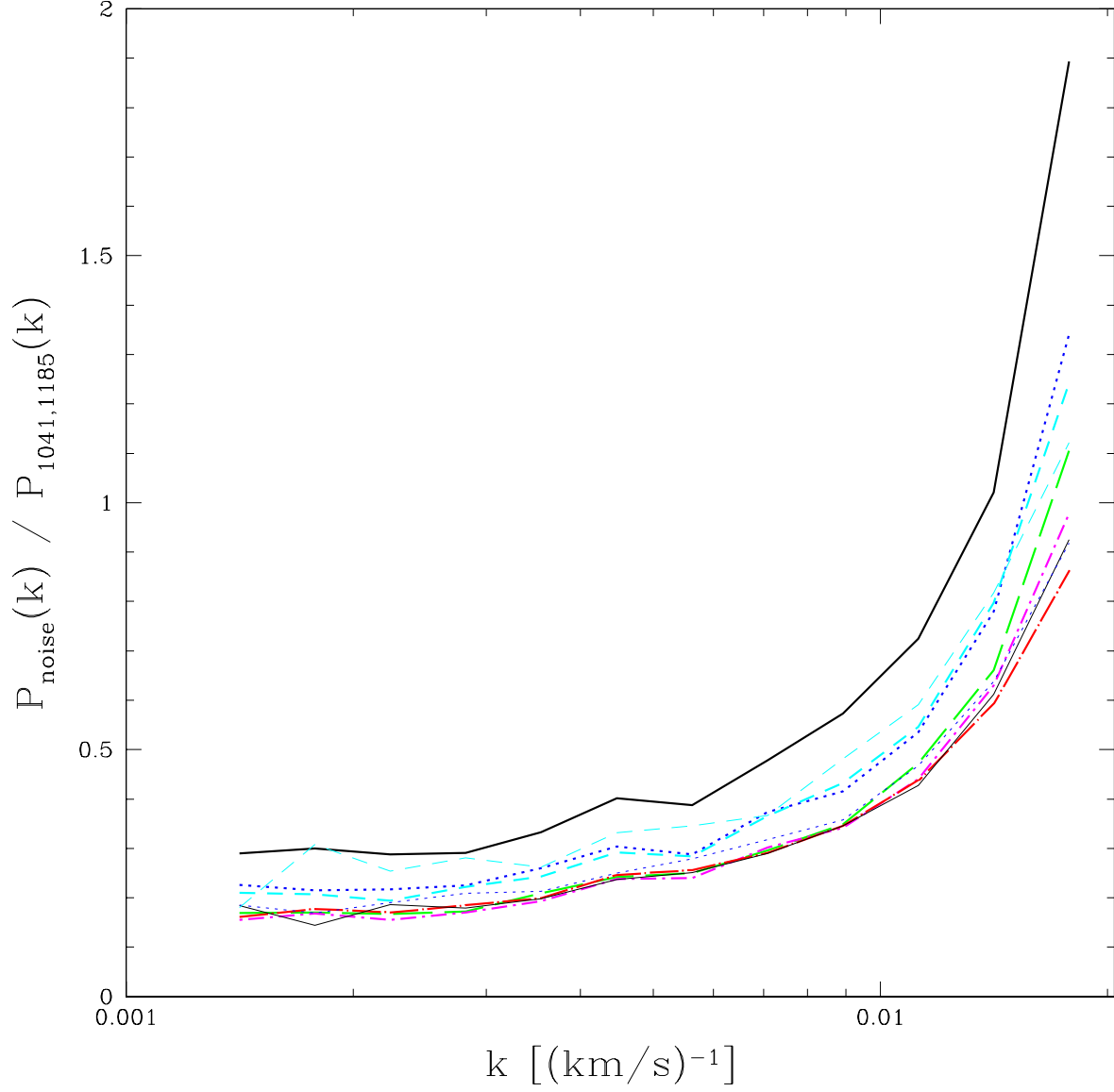


Fig. 11.— Lines, with types following the usual pattern (see Fig. 9), connect the quantity $P_{\text{noise}}/P_{1041,1185}$ for each measured point (the highest line is the lowest redshift).

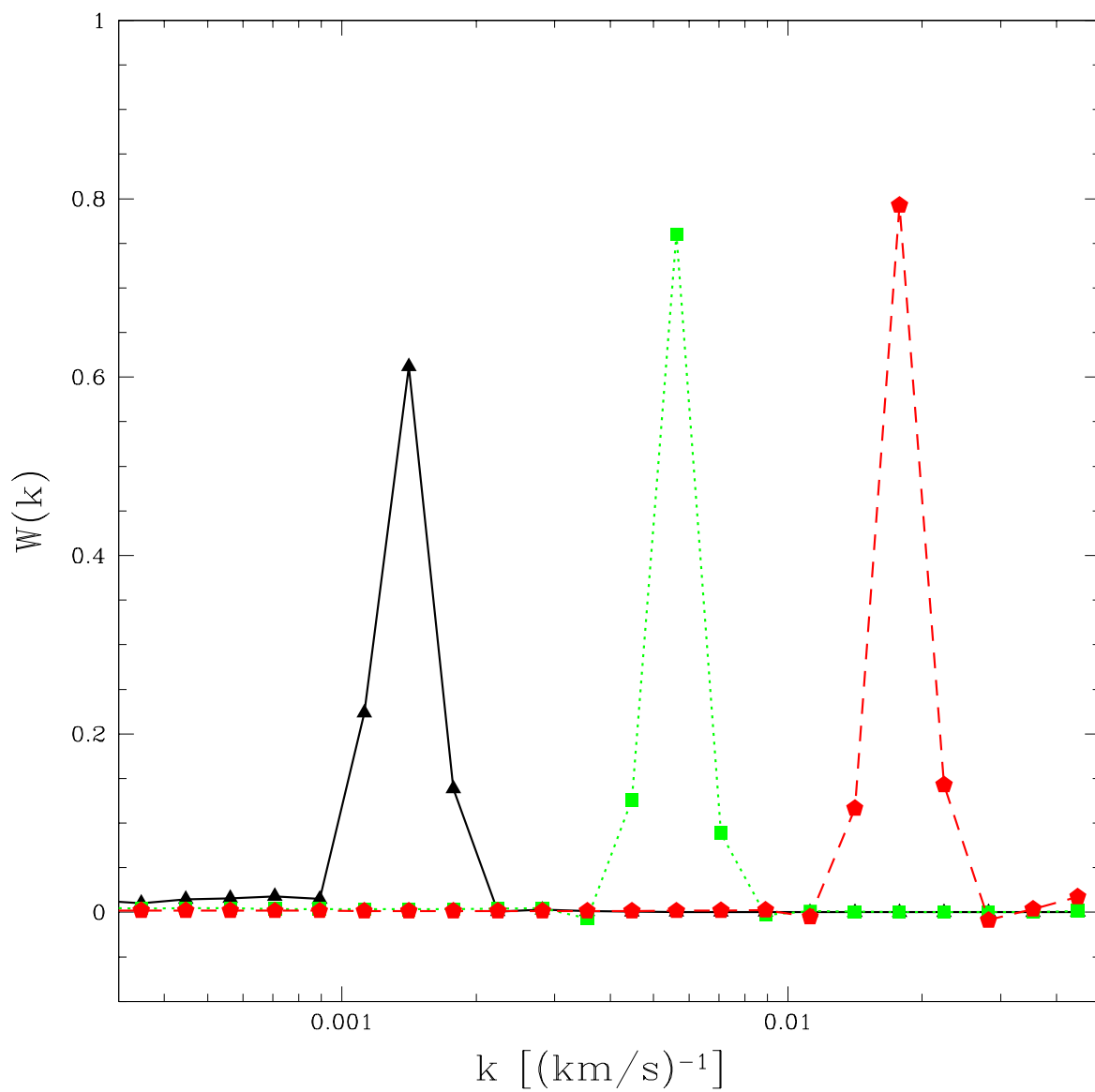


Fig. 12.— The window matrix for bands indicated by the maximum (before diagonalization).

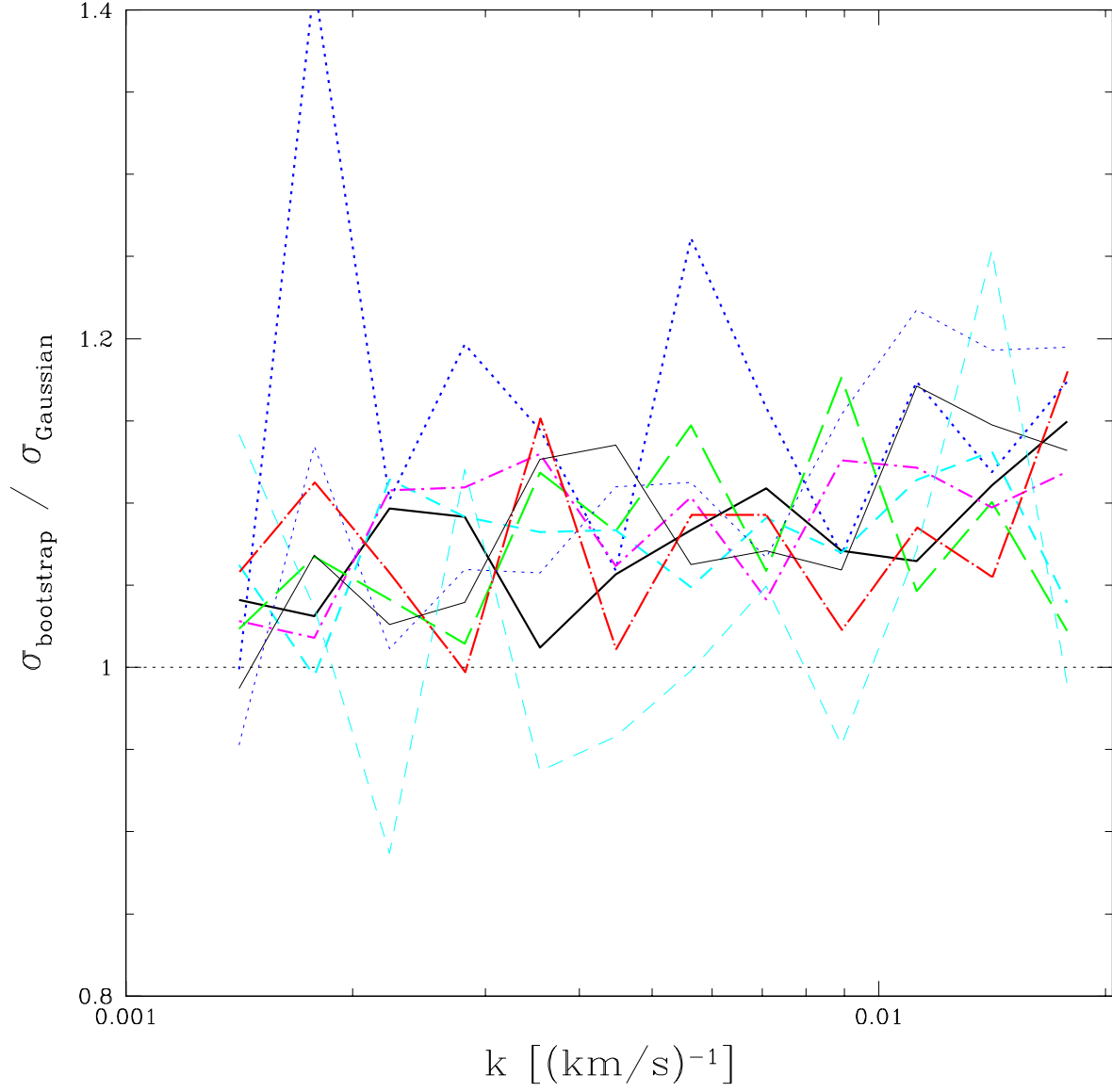


Fig. 13.— The ratio of bootstrap errors to the Gaussian estimate of the errors. See Fig. 9 for the correspondence between lines and redshift bins.

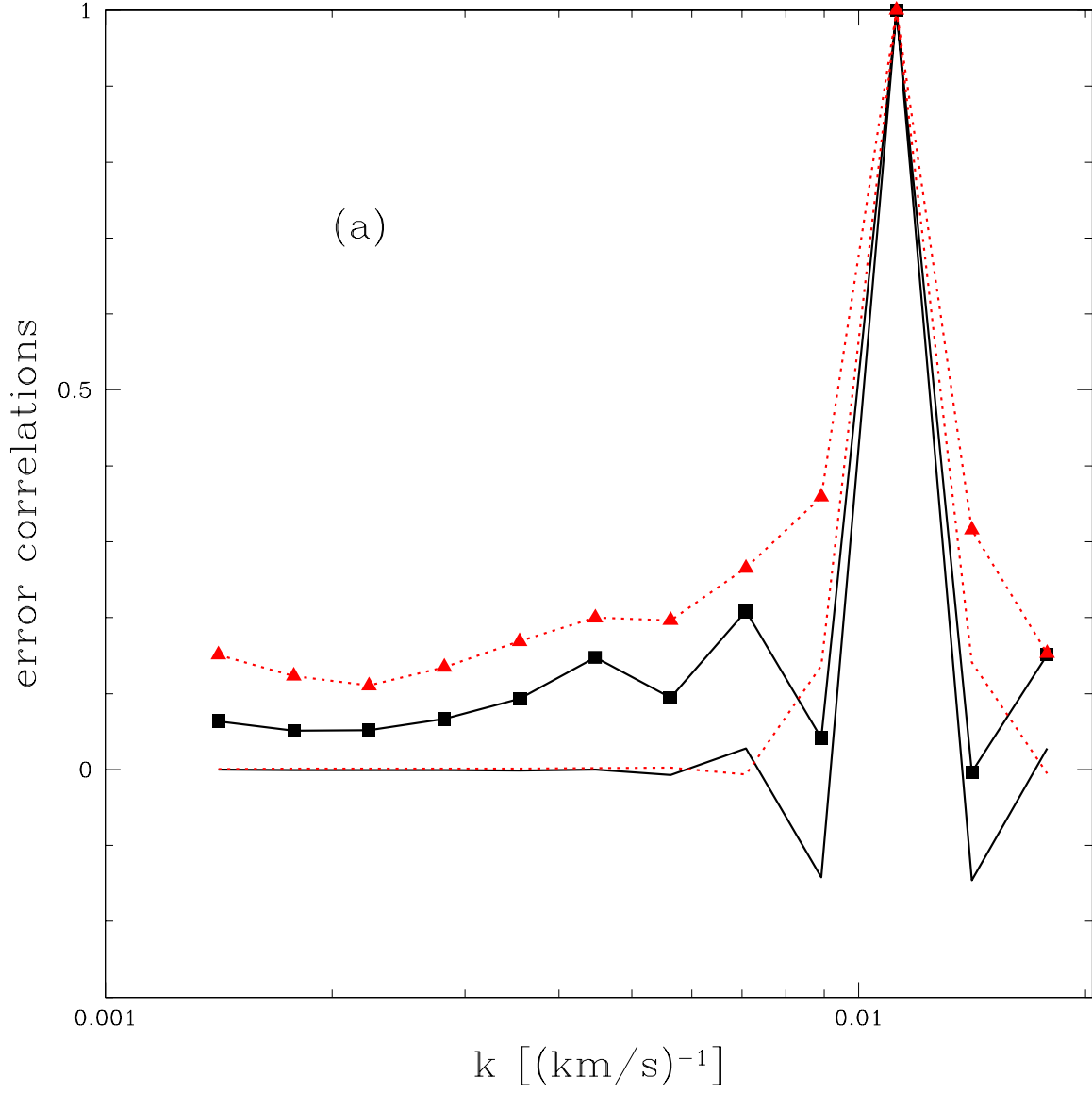
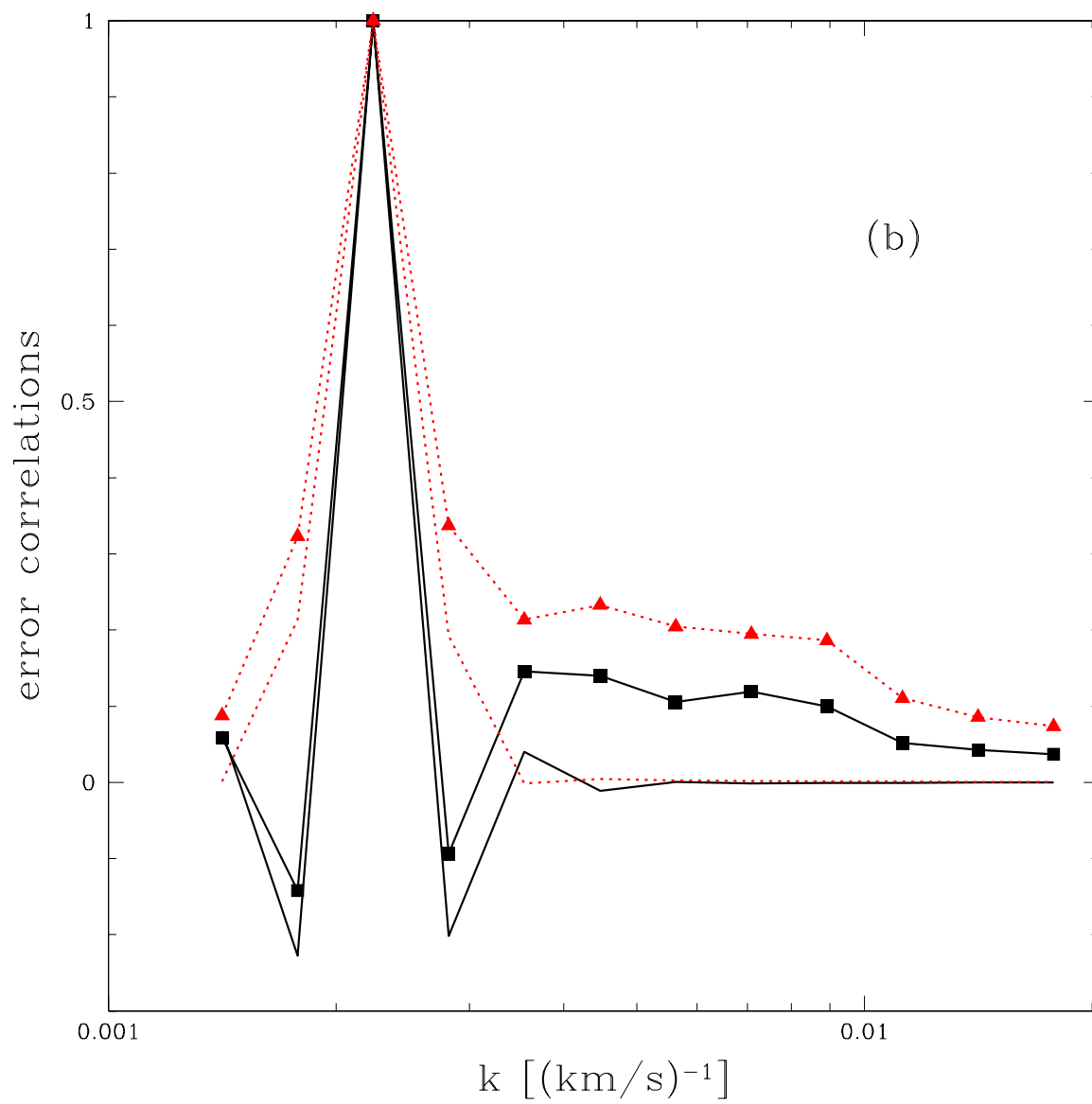


Fig. 14.— Examples of the correlations between the errors, $\langle \Delta P_i \Delta P_j \rangle / \sigma_{P,i} \sigma_{P,j}$. The black solid lines and squares show the error correlation when the window matrix is diagonalized. The red dotted lines and triangles show the correlations between points before diagonalization. The lines marked by symbols are the bootstrap estimate, while the unmarked lines are the Gaussian estimate.



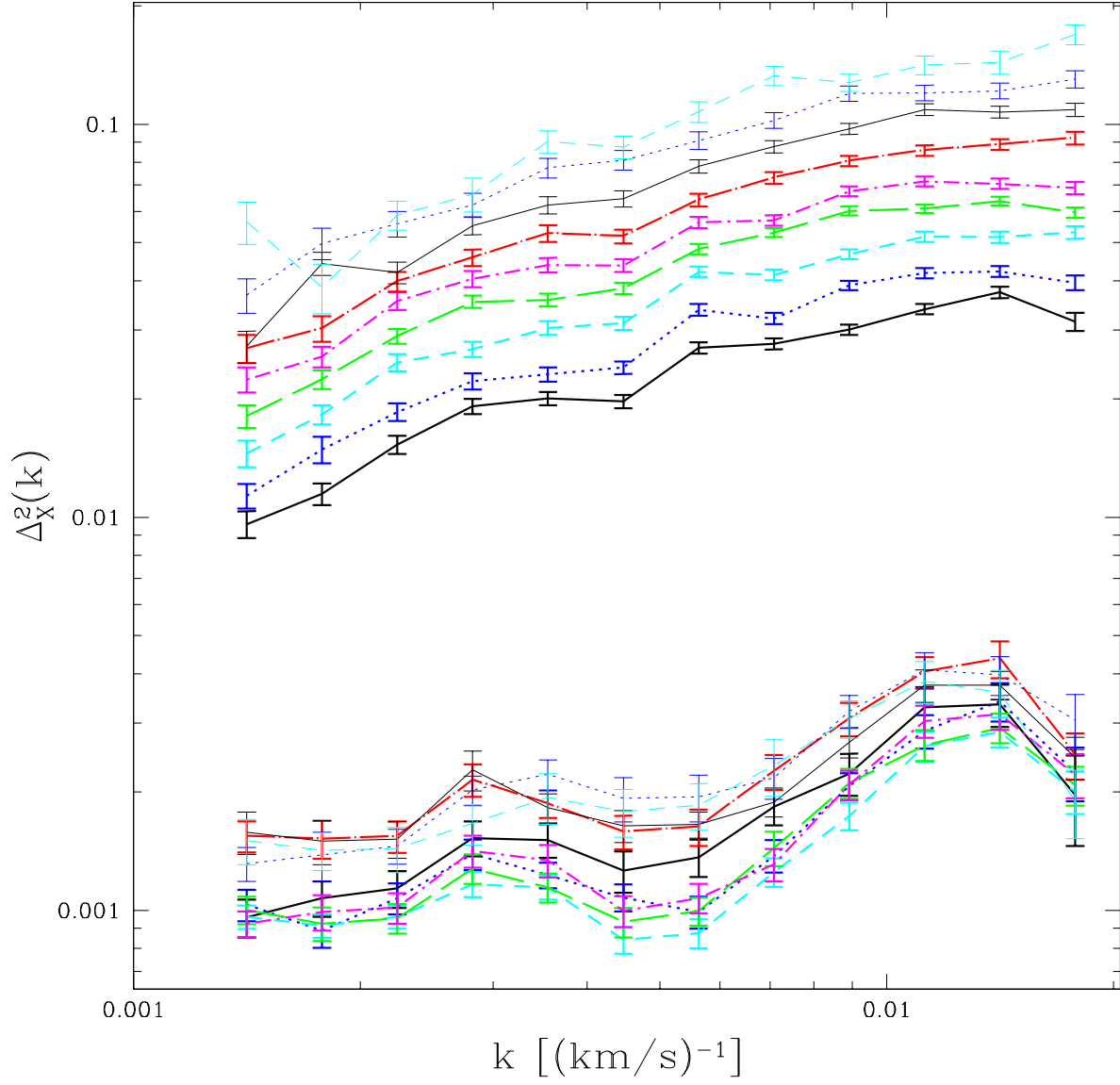


Fig. 15.— The upper set of lines show $P_{1041,1185}$, the lower set of lines show $P_{1268,1380}$. The colors and line types identify redshift bins as defined in Fig. 9.

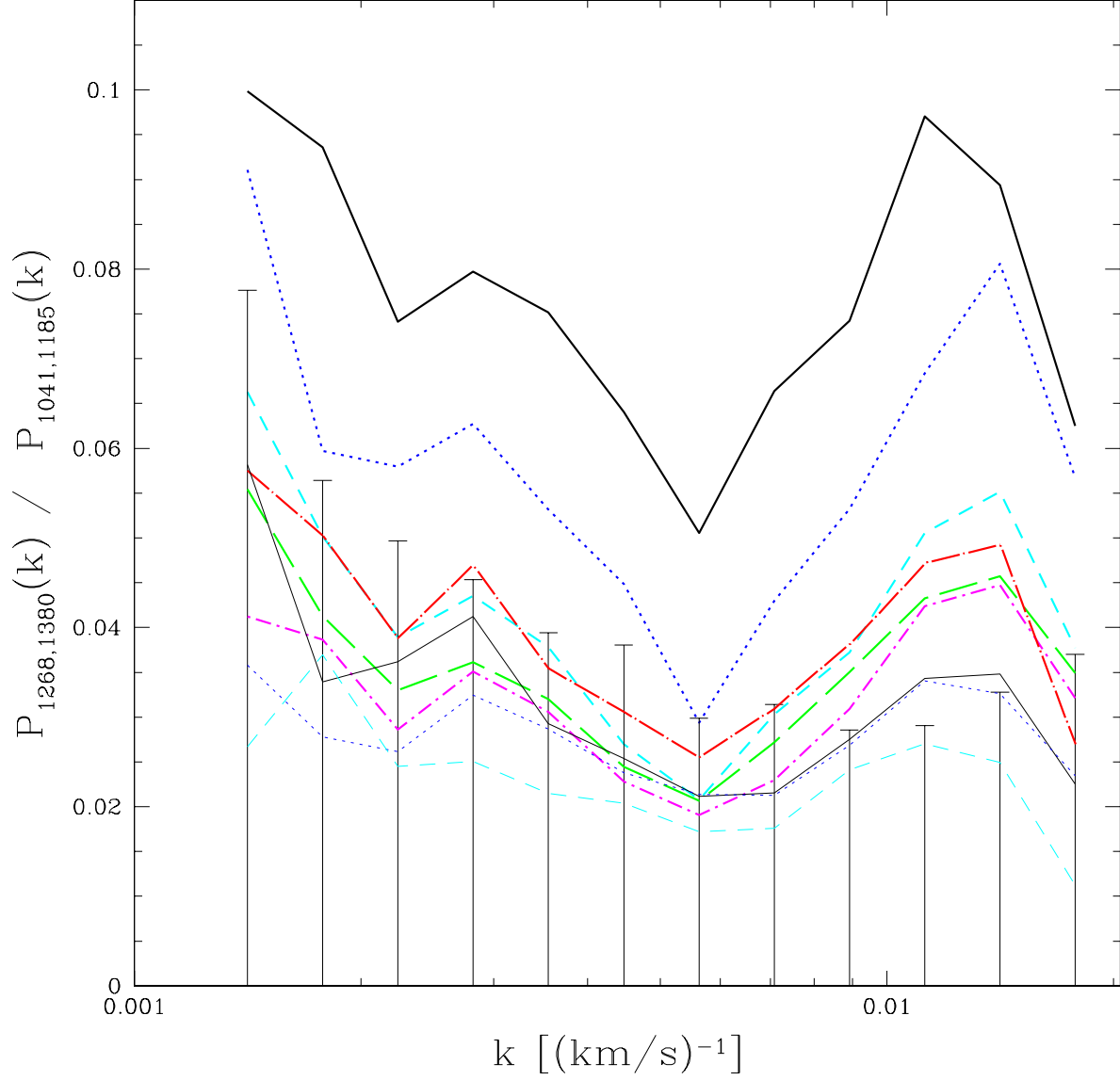


Fig. 16.— The lines show the ratio $P_{1268,1380}/P_{1041,1185}$. The uppermost (black solid) line is $z = 2.2$, and the next (blue dotted) is $z = 2.4$ (see Fig. 9 for the rest of the line definitions). For reference, the error bars starting at zero show the fractional errors on $P_{1041,1185}(k, z = 2.6)$, which are much larger than the errors on $P_{1268,1380}$ (we are simply plotting $\sigma_P(k)/P(k)$ as in Fig. 10, except that, for clarity, we show error bars starting at zero instead of a connected line).

many data points (e.g., a $1/2 \sigma$ error over 100 points shifts the mean by 5σ).

We are going to subtract the power in the range 1268-1380 Å from the Ly α forest power, but it is informative to measure the power at other places in the quasar rest frame for comparison. The range 1409-1523 Å includes CIV absorption (at 1548.2 and 1550.78 Å) but excludes SiIV (at 1393.75 and 1402.77 Å) and shorter wavelength transitions. Figure 17 shows $P_{1409,1523}/P_{1041,1185}$. If all of the power was coming from metal line absorption, the power in the range $1409 < \lambda_{\text{rest}} < 1523$ Å should always be less than the power in the range $1268 < \lambda_{\text{rest}} < 1380$ Å. As we see in Figure 18, which shows the difference in the background fractions, $(P_{1268,1380} - P_{1409,1523})/P_{1041,1185}$, the power in $P_{1268,1380}$ is greater than $P_{1409,1523}$ except on large scales. The difference on large scales suggests that there is tiny amount of power left in the quasar continua (in spite of our division by the mean continuum), which is larger in the range 1409-1523 Å than in the range 1268-1380 Å. Finally, Figure 19 shows $P_{1558,1774}/P_{1041,1185}$, past the wavelength of CIV absorption. The reduction of power relative to shorter wavelengths is dramatic, but not surprising since CIV is the most common metal absorber. It does suggest however that most of the power is due to metals and not continuum fluctuations, unless the continuum in the range 1558-1774 Å has significantly less power relative to other intervals studied here (which is admittedly not inconceivable). It seems likely, although we are not certain, that the $z = 2.2$ background power has a noticeable contribution from measurement-related problems, because the alternative is a very sudden increase in metal absorption power.

What is the upshot from these studies? The metal absorption appears to contribute a small but significant amount of power, which should also appear in the Ly α forest region. We subtract this power from the power measured in the forest. There is some indication of measurement-related problems in our lowest redshift bin. The power contributed by deviations of the quasar continua from their mean appears to be small.

While the idea that $P_{1268,1380}$ contains almost exclusively power due to simple metal absorption seems plausible at this point, when we perform consistency checks in §4.4 we find evidence that this is not the case. Splitting the data set used to measure $P_{1268,1380}$ in half based on the noise level in each spectrum, we find that the power in the halves is significantly different, by as much as 50% in some bins. Splits based on several other properties of the data (e.g., sky to quasar flux ratio) also show significant differences, but we find that these differences can all be accounted for by the difference in the basic noise level in the subsamples. Splits of the Ly α forest data set show similar trends in $P_{1041,1185}$ with the splitting parameters, although the fractional differences are much smaller. While we don't know the source of this noise dependence, it is not hard to imagine relatively benign reasons for it. For example, if sky subtraction is imperfect this would add an increasing amount of power as the sky flux, and thus noise level, increases relative to the quasar flux. The procedure we describe next would remove this power.

Since we know that $P_{1268,1380}$ depends on noise it seems logical to subtract the value of $P_{1268,1380}$ corresponding to the level of noise in the forest, rather than the best measured value of $P_{1268,1380}$,

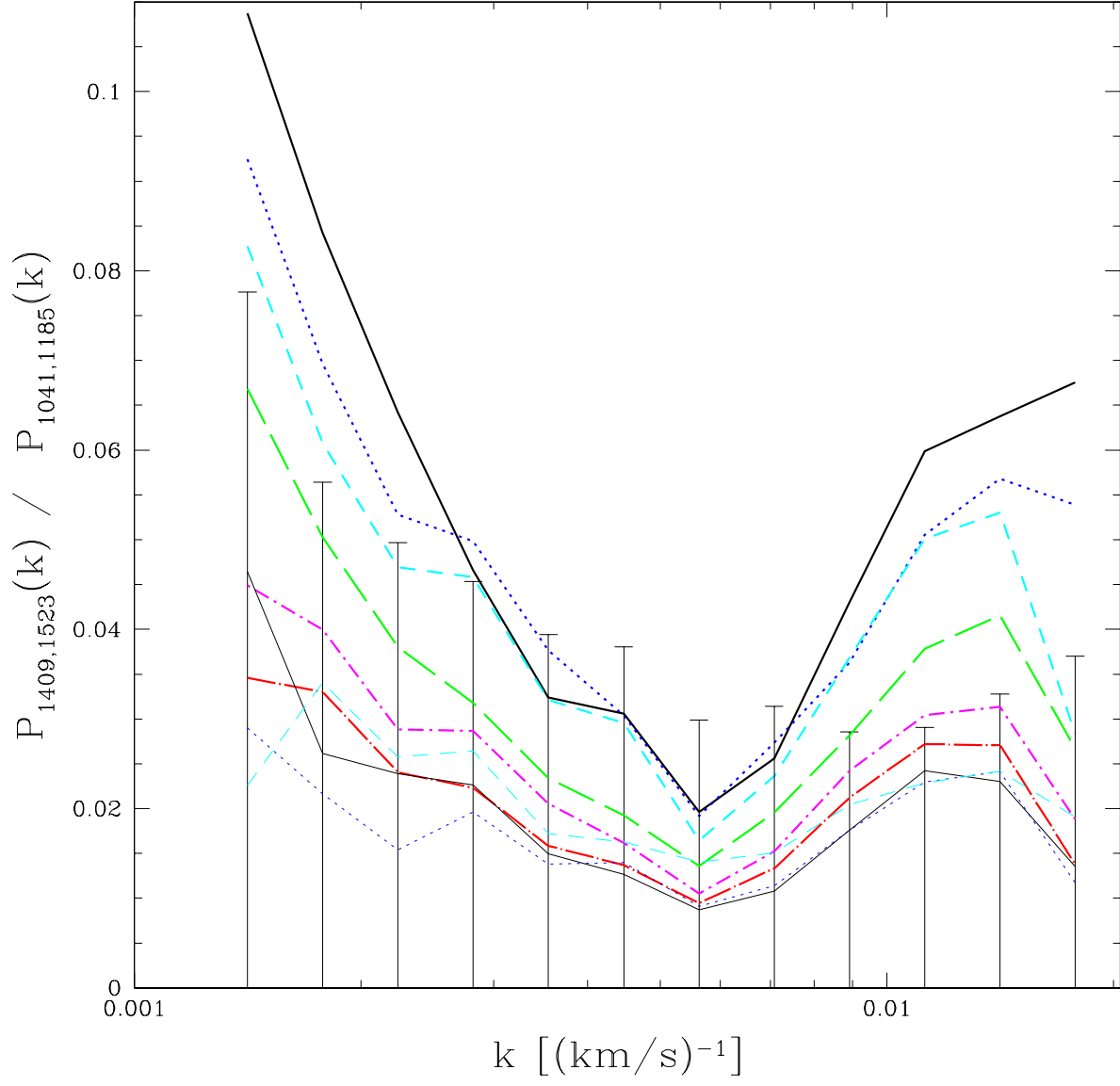


Fig. 17.— The ratio of $P_{1409,1523}$ to $P_{1041,1185}$.

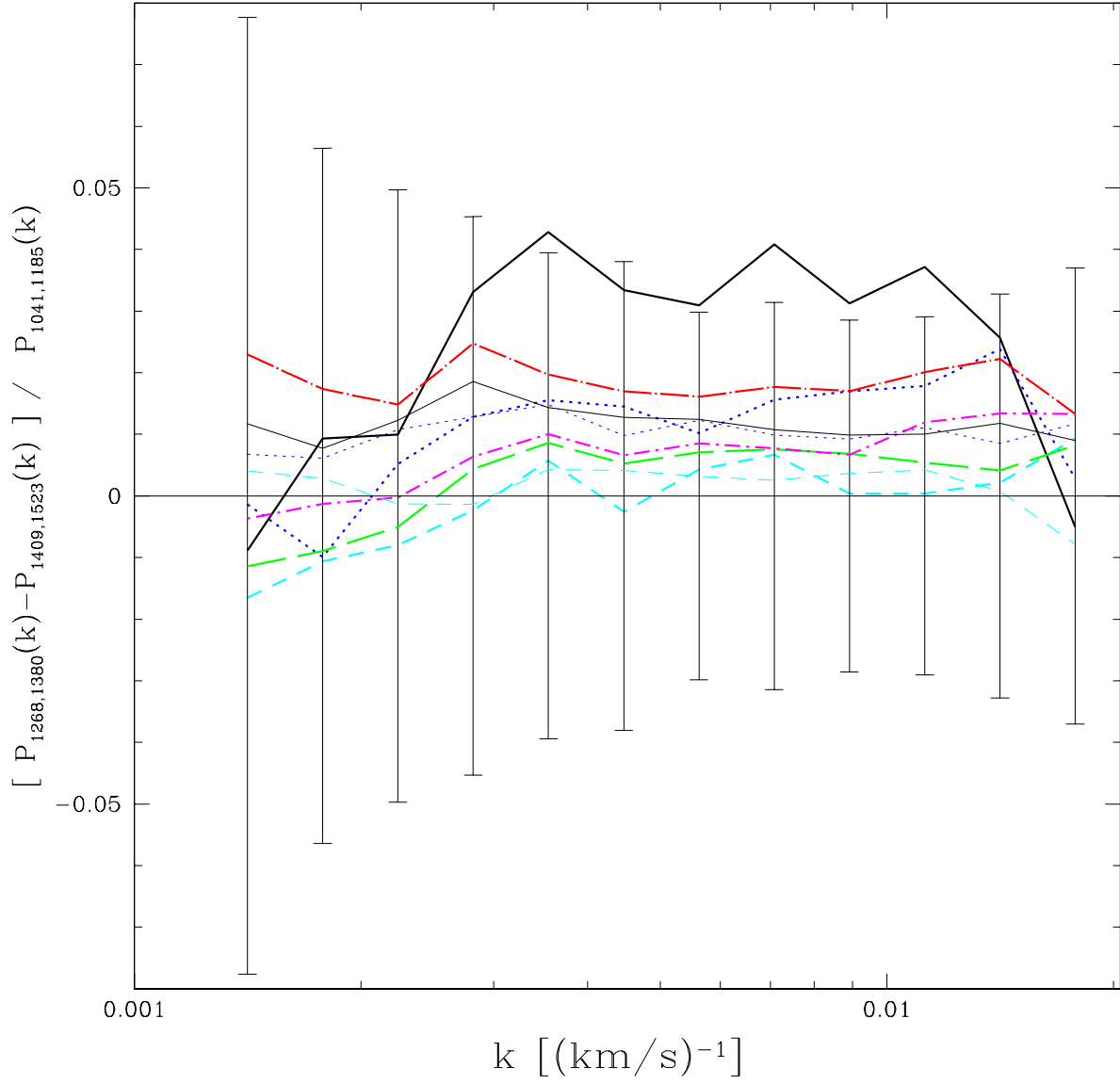


Fig. 18.— The difference between $P_{1268,1380}$ and $P_{1409,1523}$, divided by $P_{1041,1185}$, with the fractional errors on $P_{1041,1185}(k, z = 2.6)$ plotted as usual.

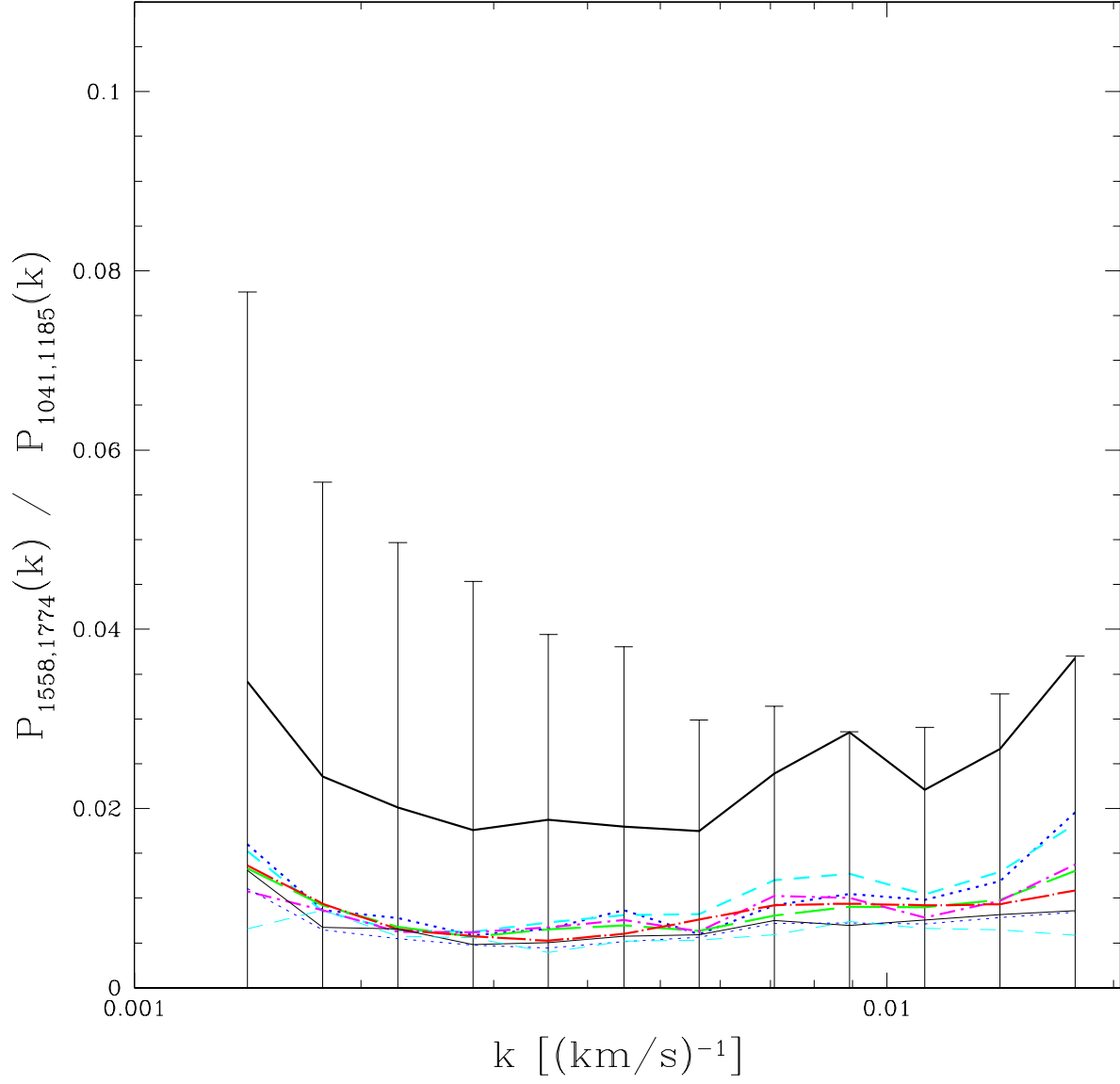


Fig. 19.— The ratio of $P_{1558,1774}$ to $P_{1041,1185}$. The uppermost (black solid) line is $z = 2.2$.

which is dominated in practice by data with less noise. If we had simply misestimated the noise by an overall factor, for example, the errors in $P_{1041,1185}$ and $P_{1268,1380}$ would cancel for this form of subtraction. To implement this idea, we model the background subtraction term as a linear function of the noise level,

$$P_{1268,1380}(k, z, \bar{\sigma}_w) = A(k, z) + B(k, z) \bar{\sigma}_w , \quad (13)$$

where $\bar{\sigma}_w$ is the mean noise level in the data computed in the same way as the mean flux level (this is the mean of the normalized noise, i.e., after division by continuum and mean flux). The choice of a linear relation is arbitrary but it does the job (see §4.4) better than the alternatives we tried. We compute $A(k, z)$ and $B(k, z)$ for each value of k and z using a linear fit to the full sample of spectra that probe $P_{1268,1380}(k, z)$, weighted by the Gaussian estimate of the error on each point for each spectrum. When the time comes to subtract the background from $P_{1041,1185}$ to obtain P_F , we use $\bar{\sigma}_w$ computed in the 1041 – 1185Å wavelength range to compute the appropriate subtraction term. Figure 20 shows the extra power subtracted through Equation (13), beyond what we would subtract if we simply used $P_{1268,1380}(k, z)$ from Figure 16. It is typically less than 4% of the Ly α forest power, but rises to 10% at the highest k for the lowest redshift.

The reader who is paying attention may complain that we have no compelling reason to believe that this source of noise-dependent power that we do not understand depends on noise in the same way inside and outside the Ly α forest region. This would be true, except that when we follow this prescription for background subtraction the differences in $P_{1041,1185}$ between subsamples are removed (see §4.4). This would be a remarkable coincidence if our model for the subtraction was not substantially correct.

Note that the background power has much smaller (absolute) statistical errors than $P_{1041,1185}$, mostly because there is simply less power, but also because there are more quasars probing a given redshift interval.

4. Consistency Checks

In §4.1 we describe how we use fits of theoretical models to the $P_F(k, z)$ results to help understand the importance of any systematic errors. We plot the correlation function of the Ly α forest in §4.2 and use it to identify a significant contribution to $P_F(k, z)$ from SiIII absorption. In §4.3 we examine the effects of modifications of our procedure. In §4.4 we break the data set up in many different ways to look for dependencies of $P_F(k, z)$ that should not exist. In §4.5 we discuss the possibility that continuum fluctuations contribute significant power. Finally, we compare our results to past measurements in §4.6.

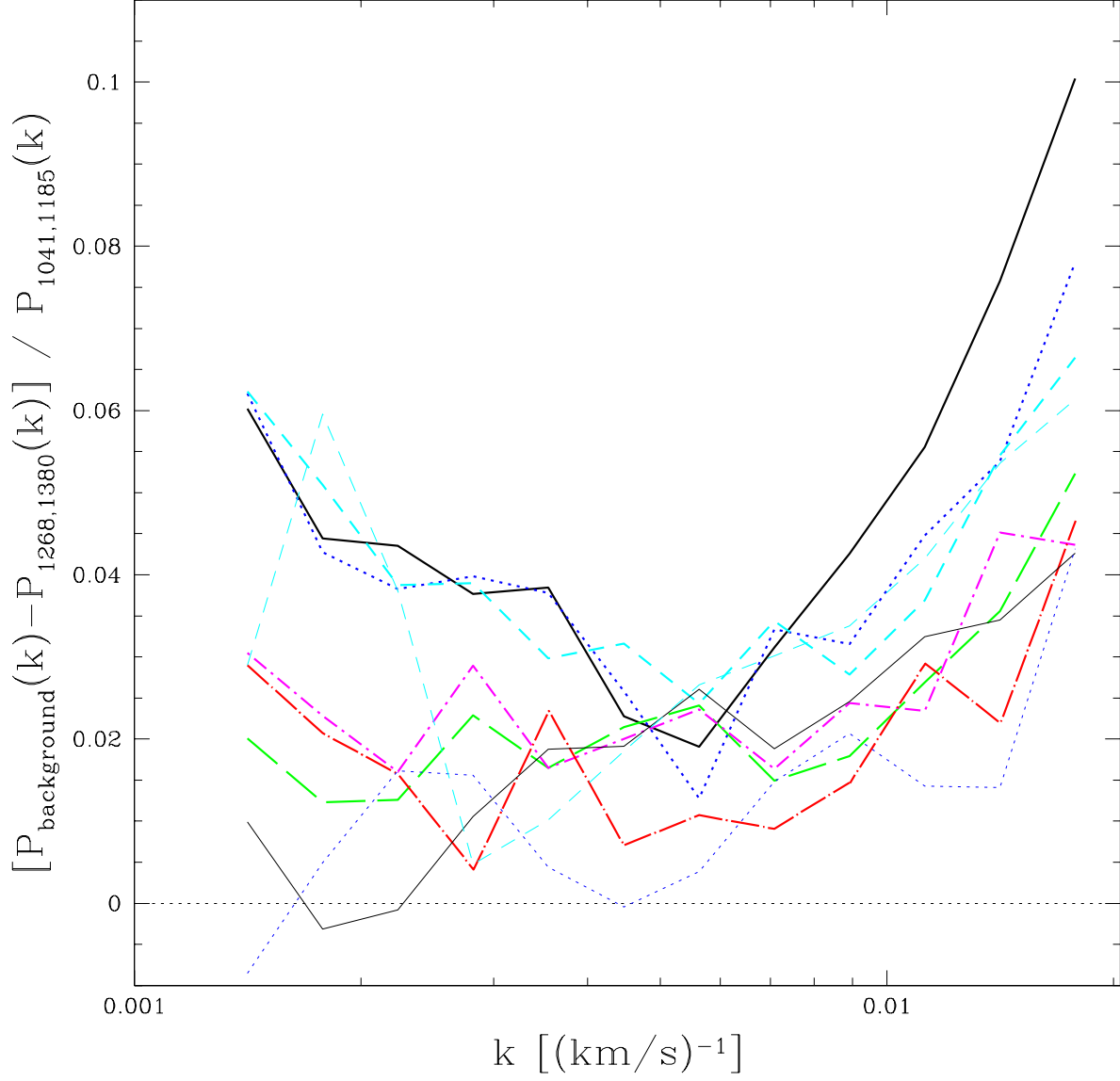


Fig. 20.— The difference between the noise dependent background power that we subtract through Equation (13) and $P_{1268,1380}$, relative to $P_{1041,1185}$, i.e., this is the extra fractional power that is subtracted because we correct for the difference between the typical noise level in the forest and in the range 1268-1380 Å.

4.1. Rudimentary Fitting of Theoretical Models

The ultimate purpose of measuring the Ly α forest power spectrum is to determine cosmological parameters by comparing the observed $P_F(k, z)$ to the predictions for different cosmological models. For the Λ CDM models supported by present observations, the universe is nearly Einstein-de Sitter at $z > 2$, so cosmology influences the Ly α forest almost entirely through the linear theory power spectrum of the mass fluctuations, $P_L(k, z)$, at $z \sim 3$ and $k \sim 0.01 \text{ (km s}^{-1}\text{)}^{-1}$ (roughly 1 comoving h/Mpc, depending somewhat on the model). We usually parameterize $P_L(k, z)$ by its amplitude, $\Delta^2(k_p, z_p) \equiv k_p^3 P_L(k_p, z_p)/2\pi^2$, slope $n_{\text{eff}}(k_p, z_p) \equiv d \ln P_L / d \ln k|_{k_p, z_p}$, and curvature $\alpha_{\text{eff}}(k_p, z_p) \equiv dn_{\text{eff}} / d \ln k|_{k_p, z_p}$, where we use $k_p = 0.009 \text{ (km s}^{-1}\text{)}^{-1}$ and $z_p = 2.6$ as the pivot points.

A full discussion of the details of theoretical modeling of $P_F(k, z)$ using numerical simulations is beyond the scope of this paper. Furthermore, the theory of the Ly α forest is perhaps less certain than the observations, so we want to present the observational results un-tarnished by theoretical interpretation. However, it is very useful to interpret the possible systematic errors in the appropriate context of cosmological model fitting. In other words: without model fitting, it is difficult to know how important a given change in $P_F(k, z)$ is.

In this paper we take a cautious approach to the theoretical model fitting – we perform fits to different estimates of $P_F(k, z)$ computed using modifications of the extraction procedure or different subsamples of the data, however, we do not give the central result, only the deviations in the results from the value obtained from our preferred $P_F(k, z)$. These deviations in fitting results should give the reader a useful indication of the importance of systematic effects in the data, regardless of the reader’s opinion of the theory.

The simulations and fitting procedure that we use are described in McDonald et al. (2004), where we present the final result. We use a Λ CDM transfer function, and perform the fit with Δ^2 and n_{eff} as free parameters (because $\alpha_{\text{eff}} = dn_{\text{eff}} / d \ln k$ is not tightly constrained by the present Ly α forest data alone, we fix the primordial running $\alpha = dn / d \ln k$, not to be confused with $\alpha_{\text{eff}} \simeq -0.2$, to zero). Unless otherwise specified, we perform the fits using the 108 $P_F(k, z)$ points in the ranges $0.0013 < k < 0.02 \text{ (km s}^{-1}\text{)}^{-1}$ and $2.1 < z < 3.9$. We allow for some error in our noise estimate by permitting the noise subtraction terms to vary independently in each redshift bin by 5% (9 extra free parameters to fit for, constrained by Gaussian likelihood function with this rms width). We also allow a single overall parameter describing the squared resolution error to vary with rms constraint $(7 \text{ km s}^{-1})^2$.

The Ly α forest model in the simulations is controlled by the externally constrained functions $\bar{F}(z)$, the mean absorption, $T_{1.4}(z)$, the temperature at overdensity 1.4, $(\gamma - 1)(z)$, the logarithmic slope of the temperature-density relation, and a reionization parameter that we will call x_{rei} . $\bar{F}(z)$ is described in our fits by the 10 parameter formula $\bar{F}_i = \mathcal{F} \tilde{F}_i$, where i labels our 9 redshift bins, \tilde{F}_i gives the arbitrarily normalized z dependence and \mathcal{F} is an overall normalization. We have performed a preliminary analysis using the formalism in §2.5 to measure $\bar{F}(z)$ from SDSS data and we use this to constrain the parameters \tilde{F}_i (the error on each redshift bin is ~ 0.005). Because the SDSS

analysis does not constrain the overall normalization, we leave \mathcal{F} free except for the additional constraint that we require \bar{F}_i interpolated to $z = (3.9, 3.0, 2.4)$ to match the HIRES constraints $\bar{F} = (0.458 \pm 0.034, 0.676 \pm 0.032, 0.816 \pm 0.023)$ (see McDonald et al. (2000) – we have modified the numbers slightly and increased the errors to allow for systematic uncertainties, as discussed in Seljak et al. (2003)). We parameterize $T_{1.4}(z)$ and $(\gamma - 1)(z)$ by quadratic functions of z (3 parameters each) with the external constraints $T_{1.4} = (20100 \pm 3400, 20300 \pm 2400, 20700 \pm 2800)\text{K}$ and $\gamma - 1 = (0.43 \pm 0.45, 0.29 \pm 0.3, 0.52 \pm 0.14)$ at $z = (3.9, 3.0, 2.4)$ (see McDonald et al. (2001) – we added 2000 K in quadrature to the temperature errors to allow for systematic errors). Note that there are other, sometimes more precise, measurements of \bar{F} (Schaye et al. 2003; Bernardi et al. 2003) and the temperature-density relation (Schaye et al. 2000; Ricotti et al. 2000) in the literature – our choice of McDonald et al. (2000) and McDonald et al. (2001) for this example is simply for convenience. The redshift of reionization and post-ionization temperature of the gas influence Ly α forest predictions because the smoothing of the gas on small scales depends on its pressure history. At the level of precision we care about, this dependence can be captured by a single parameter. In our modeling, we use x_{rei} to interpolate between two reasonable boundaries, reionization heating of the gas to 25,000 K at $z = 7$ or to 50,000 K at $z = 17$, both of which are consistent with our temperature constraints $T_{1.4}(z)$. However, in this paper we fix x_{rei} , because it is weakly constrained by the data and the hard lower limit we have to impose on the redshift of reionization leads to non-Gaussian errors on the power spectrum parameters we are interested in (this is a problem of presentation, not of principle).

Figure 21 shows our first fit to the standard $P_F(k, z)$ results. The value of $\chi^2 = 193.7$ is much too high for approximately 106 degrees of freedom (we are marginalizing over a large number of nuisance parameters, but these generally are externally constrained so they do not reduce the number of degrees of freedom). Including α_{eff} as a free parameter does not improve the fit significantly. It appears that much of the disagreement comes from bumps in the power spectrum, e.g., at $k \sim 0.003 \text{ (km s}^{-1}\text{)}^{-1}$. This motivates us to look at the correlation function of the flux.

4.2. The Correlation Function and the SiIII Cross-Correlation

Sometimes features in the power spectrum are easier to understand by looking at the correlation function, $\xi(v) = \langle \delta(x)\delta(x+v) \rangle$ (v is as usual a stand-in for wavelength differences, as is x in this case). We show the normalized correlation function, $\xi(v)/\xi(0)$ for the first six redshift bins in Figure 22. The correlation function shows the expected behavior – positive for small v , negative for large v – except for an obvious bump at $v \sim 2200 \text{ km s}^{-1}$. We focus on this bump in the inset panel of Figure 22. The most likely explanation seems to be cross-correlation between Ly α and SiIII absorption. SiIII absorbs at $\lambda = 1206.50\text{\AA}$, so the SiIII absorption by gas at some point along the line of sight will appear in the spectrum separated by 2271 km/s from the Ly α absorption by the same gas. We see that the bump in $\xi(v)$ appears at this separation, and note that the features that ruined our power spectrum fit appear at the expected multiples of $k = 2\pi/2271 = 0.0028 \text{ (km s}^{-1}\text{)}^{-1}$.

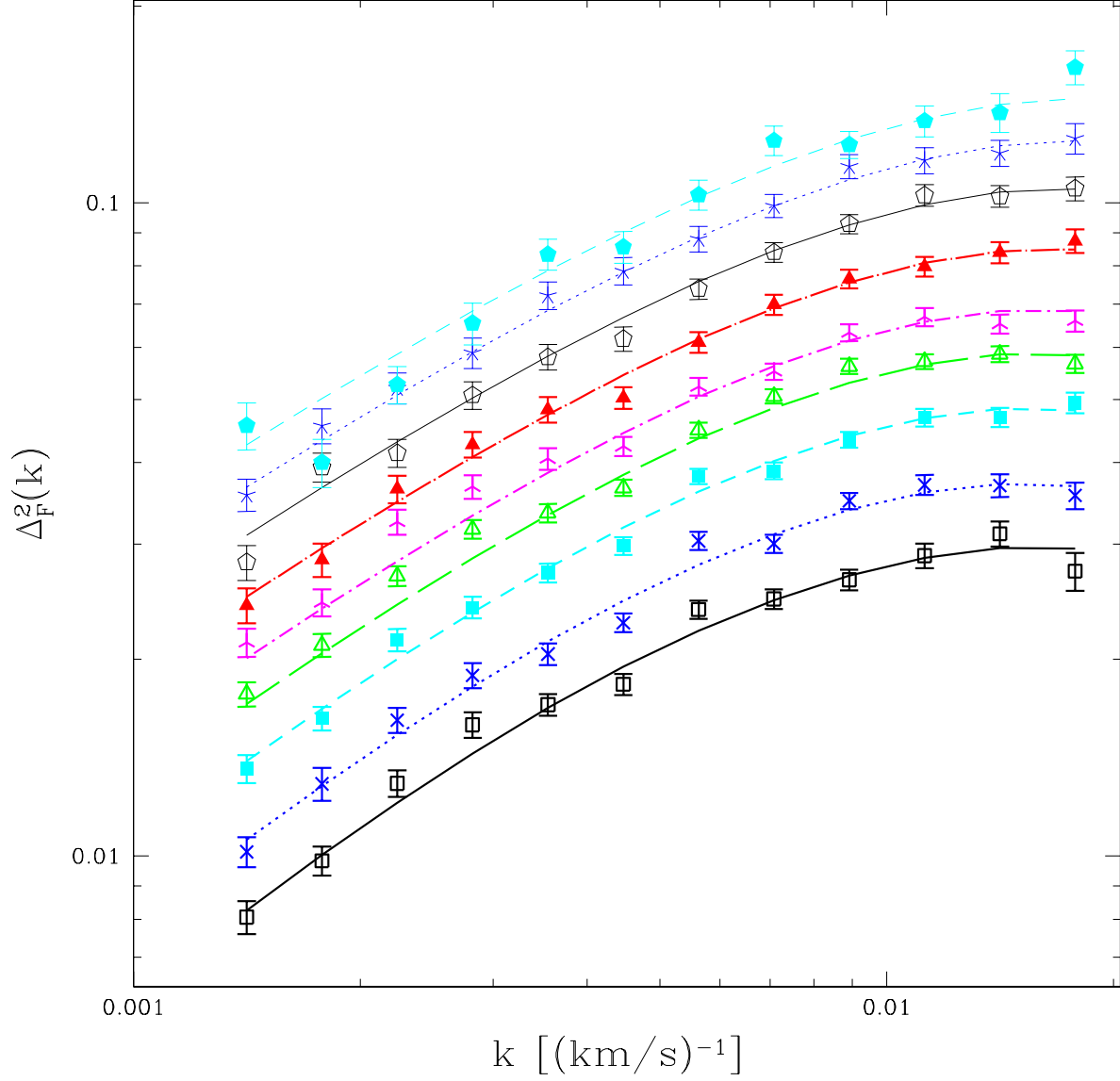


Fig. 21.— Points with error bars show the observed $P_F(k, z)$. Lines show our first attempt to fit a theoretical model, which is not a good fit to the data. From bottom to top — $z=2.2$: black, solid line, open square; $z=2.4$: blue, dotted line, 4-point star (cross); $z=2.6$: cyan, dashed line, filled square; $z=2.8$: green, long-dashed line, open triangle; $z=3.0$: magenta, dot-dashed line, 3-point star; $z=3.2$: red, dot-long-dashed line, filled triangle; $z=3.4$: black, thin solid line, open pentagon; $z=3.6$: blue, thin dotted line, 5-point star; $z=3.8$: cyan, thin dashed line, filled pentagon.

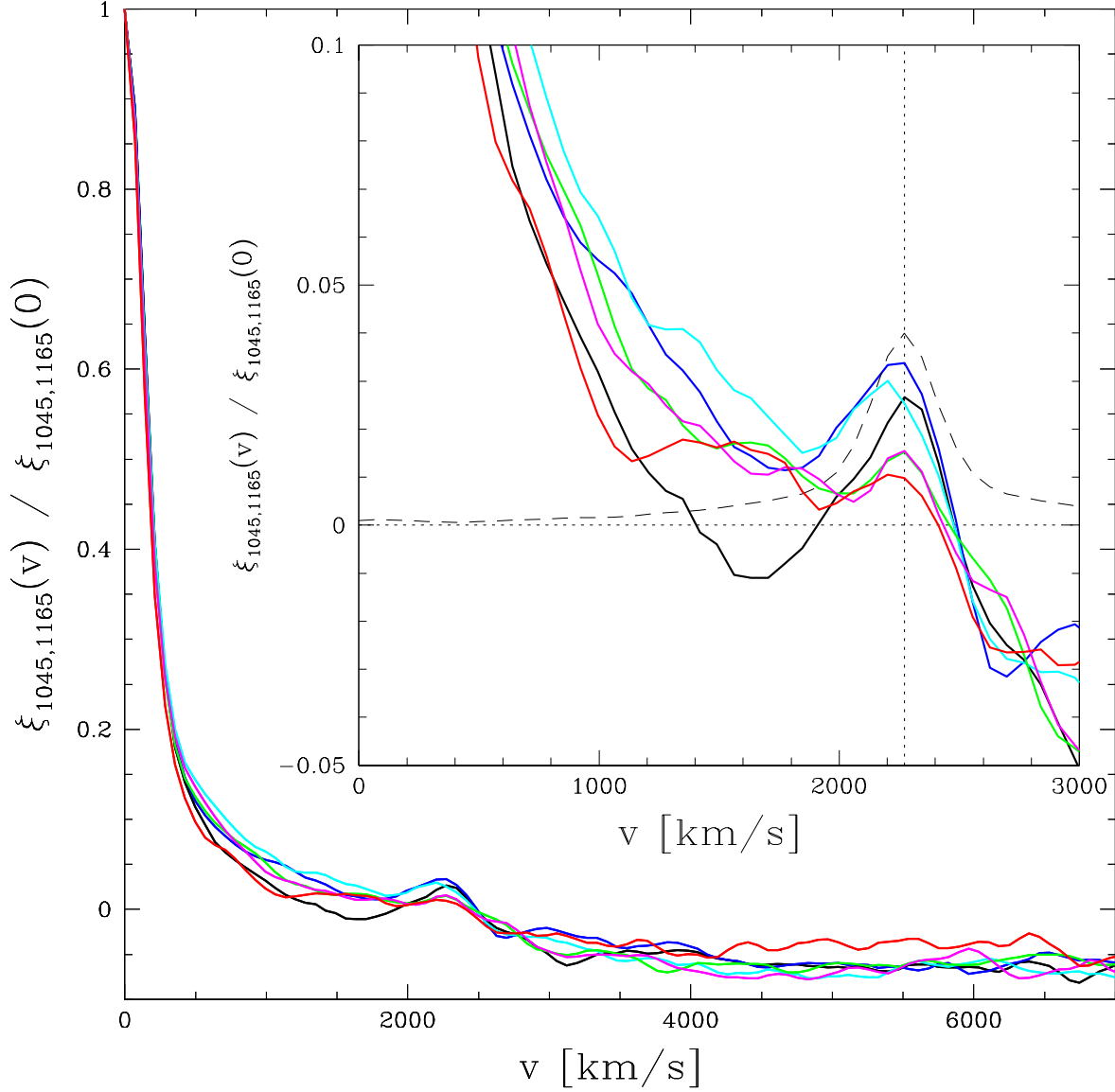


Fig. 22.— The normalized correlation function, $\xi(v)/\xi(0)$, in the Ly α forest region, uncorrected for resolution. In the inset panel, the solid lines show an expanded view of the SiIII-Ly α cross-correlation bump, the dashed line shows $0.04 \xi(v - 2271)/\xi(0)$ for comparison, and the vertical dotted line marks $v = 2271 \text{ km s}^{-1}$. Note that there is no evidence for any other metal with wavelength close to Ly α transition being important. In particular, we see no bump at $\sim 5600 \text{ km s}^{-1}$ or $\sim 6700 \text{ km s}^{-1}$, corresponding to NV-Ly α velocity differences. This correlation function should not be used for any quantitative science, as we have not corrected for resolution effects, have not checked carefully for systematic errors, and have not given statistical errors.

Note that this is the only correlation seen; another metal correlation one might expect to see is NV ($\lambda = 1238.8, 1242.8\text{\AA}$), but there is no apparent feature at the corresponding velocity differences ($\Delta v \sim 5600, 6700\text{km/s}$), as seen in figure 22.

What should we do about this SiIII-Ly α cross-correlation, since the poor χ^2 suggests that it is too significant to ignore? Our first guess might be that the SiIII-Ly α correlation is a simple offset version of the Ly α -Ly α correlation, i.e., something like $\xi_{\text{SiIII-Ly}\alpha}(v) \propto \xi_{\text{Ly}\alpha\text{-Ly}\alpha}(|v| - 2271\text{ km s}^{-1})$. The simplest way to model this is to assume that the SiIII structure is equal to that of the Ly α forest up to an overall normalization, $\delta_F = \delta(v) + a \delta(v + v_3)$, where $\delta(v)$ is for Ly α only and $v_3 = 2271\text{ km s}^{-1}$. The corresponding correlation function is

$$\xi_F(v) = (1 + a^2) \xi(v) + a \xi(v + v_3) + a \xi(v - v_3), \quad (14)$$

with corresponding power spectrum

$$P_F(k) = (1 + a^2) P(k) + 2 a \cos(v_3 k) P(k), \quad (15)$$

where unsubscripted ξ and P are understood to mean Ly α -Ly α . For our first fit to $P_F(k, z)$ accounting for SiIII using equation (15), we assume $a = f/[1 - \bar{F}(z)]$, with f as a single extra free parameter of the fit. We find a remarkable improvement in χ^2 , from 193.7 to 130.9. We find $f \sim 0.011$ ($a \sim 0.04$, depending on the redshift). The small value suggests that the relative contribution of SiIII to the autocorrelation is $a^2 < 0.004$, which will not affect our fit results significantly (see §4.3). We thus only need to estimate the cross-correlation term. We also tried allowing a power law $1 + z$ dependence for f , but the improvement in fit, $\Delta\chi^2 = 1.1$, was not significant.

In the inset panel of Figure 22 we plot scaled $\xi(v - 2271)$, to show how the shape of the bump compares to the zero-lag correlation. It is difficult to compare the shapes by eye, because of the slope of the underlying correlation, but it appears that this model explains the cross-correlation reasonably well. We can allow for a change in scale using the slightly more general form

$$\xi_F(v) = \xi(v) + a \xi[s(v + v_3)] + a \xi[s(v - v_3)]. \quad (16)$$

Allowing s to vary freely only improves χ^2 by 0.7 (note that the logarithmic k -binning that we use may reduce our ability to constrain these parameters). The error bars on other parameters may increase when we include z dependence of f and allow s to be free, so to be conservative one probably wants to leave them free even though they are not needed. In our standard fit in this paper, we allow f to have z dependence but fix $s = 1$. We show the improved fit to $P_F(k, z)$ in Figure 23.

4.3. Modifications of the Procedure

The pipeline developed for this analysis includes many improvements and corrections that were added throughout the development. It is worth taking a step back to ask how important the various

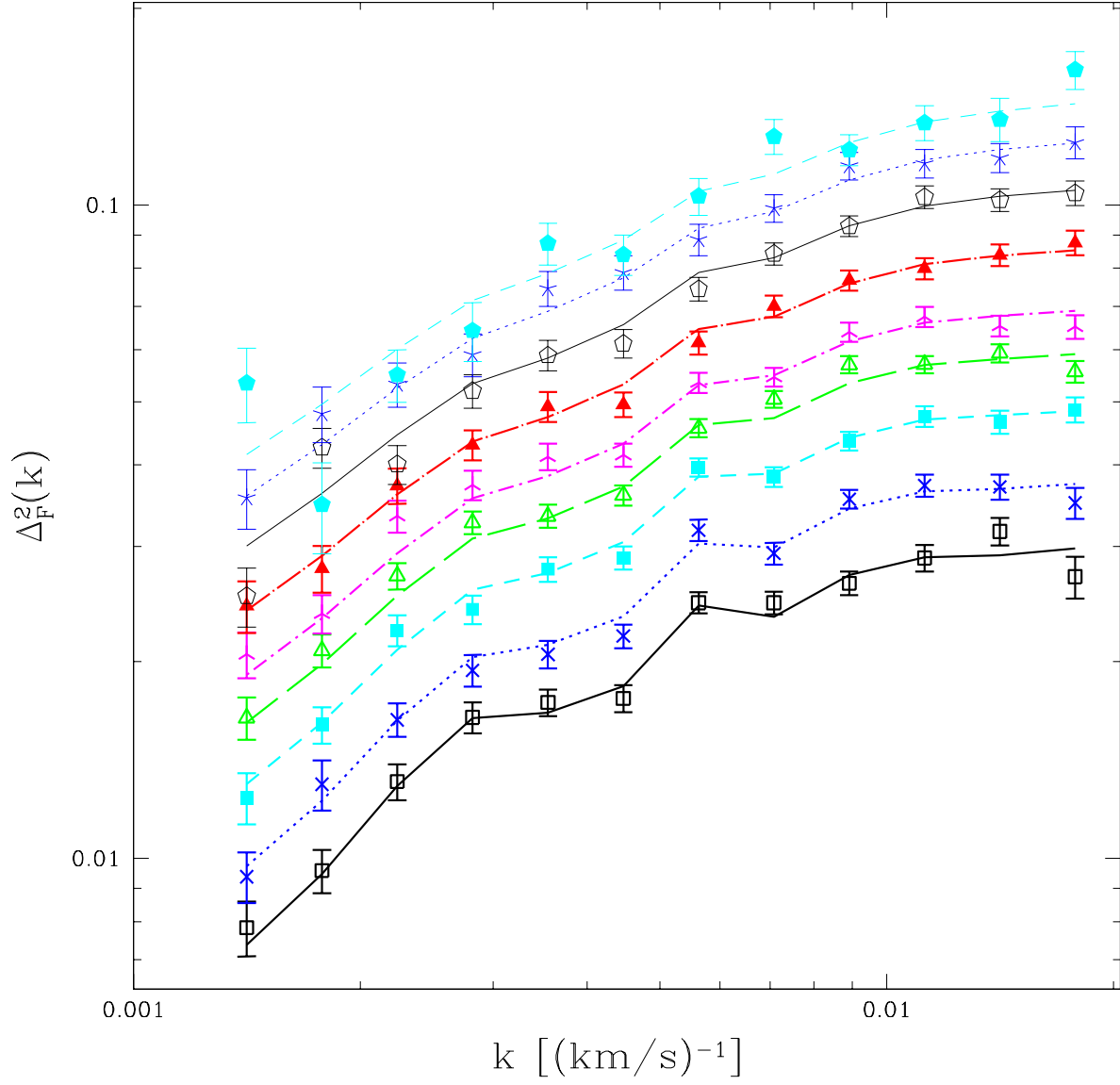


Fig. 23.— Points with error bars show the observed $P_F(k, z)$. Lines show our best fit after including SiIII absorption approximately in the theory. See Fig. 21 for line and point definitions.

corrections are for the final result. Table 1 lists various modifications of our procedure (described individually below), and quantifies their effects on the fit results. In each case we show the change in the best fit Δ^2 and n_{eff} relative to our standard fit, and their error bars for comparison to the standard fit. We give χ^2 to indicate the goodness of fit of the theory to the modified measured $P_F(k, z)$. We reiterate that we are not asserting the correctness of the theory that we use in the fitting – we give these χ^2 values and other fitting results only to show trends. We list $\Delta\chi^2$ between the standard procedure best fit and the variant best fit, to give an indication of how significant the deviation is (this is necessary because the errors are correlated so simply knowing Δ^2 and n_{eff} and their errors does not give the full picture – see Figure 26 for an example of the full error contours). Because the statistical fluctuations between these different fits should be small, a 1σ difference (or even less) should be interpreted as “significant”, however, since we believe that our standard fit is better or more conservative than all of the variants, our systematic error should generally be smaller than the deviation shown. Note that where applicable the changes in procedure are only applied to the $P_{1041,1185}$ calculation, not the $P_{1268,1380}$ result that is used in the background subtraction (small changes in $P_{1268,1380}$ have no effect on the final results).

Our first variant is to not diagonalize the window matrix. Figure 24 shows the measured power spectrum before and after diagonalization. Figure 25 shows the ratio of the diagonal errors after diagonalization of the window matrix to before diagonalization. Not diagonalizing the window matrix does lead to a significant change in the fitted parameter values, and the error on n_{eff} decreases by about 12%. We are, in effect, using information from a wider range of scales, but this forces us to use theory results outside their range of validity (e.g., at low k we need to extrapolate beyond the size of our simulation boxes). Note that the change in error on n_{eff} , from 0.024 to 0.021 implies that we should expect a random difference between the two results with typical size $(0.024^2 - 0.021^2)^{1/2} = 0.012$, i.e., what might seem like a surprisingly large part of the difference between the results could be random.

As discussed above, the correction for SiIII-Ly α correlation is very important to the goodness of our fit. It is less important for the best fit values, changing them only by 0.8σ for n_{eff} and 0.4σ for Δ^2 . Normally we allow a power law dependence on redshift for the amplitude of the SiIII-Ly α correlation, but removing this freedom makes almost no difference. Allowing the correlation scale for the SiIII-Ly α correlation to be different than for Ly α -Ly α (freeing s in eqn. 16 – we usually fix this in this paper for technical reasons) has only a very small effect. Including the SiIII-SiIII autocorrelation term (the a^2 part of eqn. 15) in the fit has essentially no effect.

For our standard fit, we allow variation in the noise amplitude at each redshift, represented by a multiplicative parameter subject to a 5% rms Gaussian constraint. Our fitting procedure then marginalizes over this component. Reducing this constraint to 0.5% (i.e., fixing the noise) produces no change in our fit result, and does not even reduce the error bars noticeably. Leaving the noise essentially free makes a noticeable difference in the fit results, decreasing the amplitude by about $2/3\sigma$, increasing its error by 20%, and decreasing χ^2 to 123.8. Changes at this level are expected when we remove the constraints on some parameters, and do not imply that the constraints were

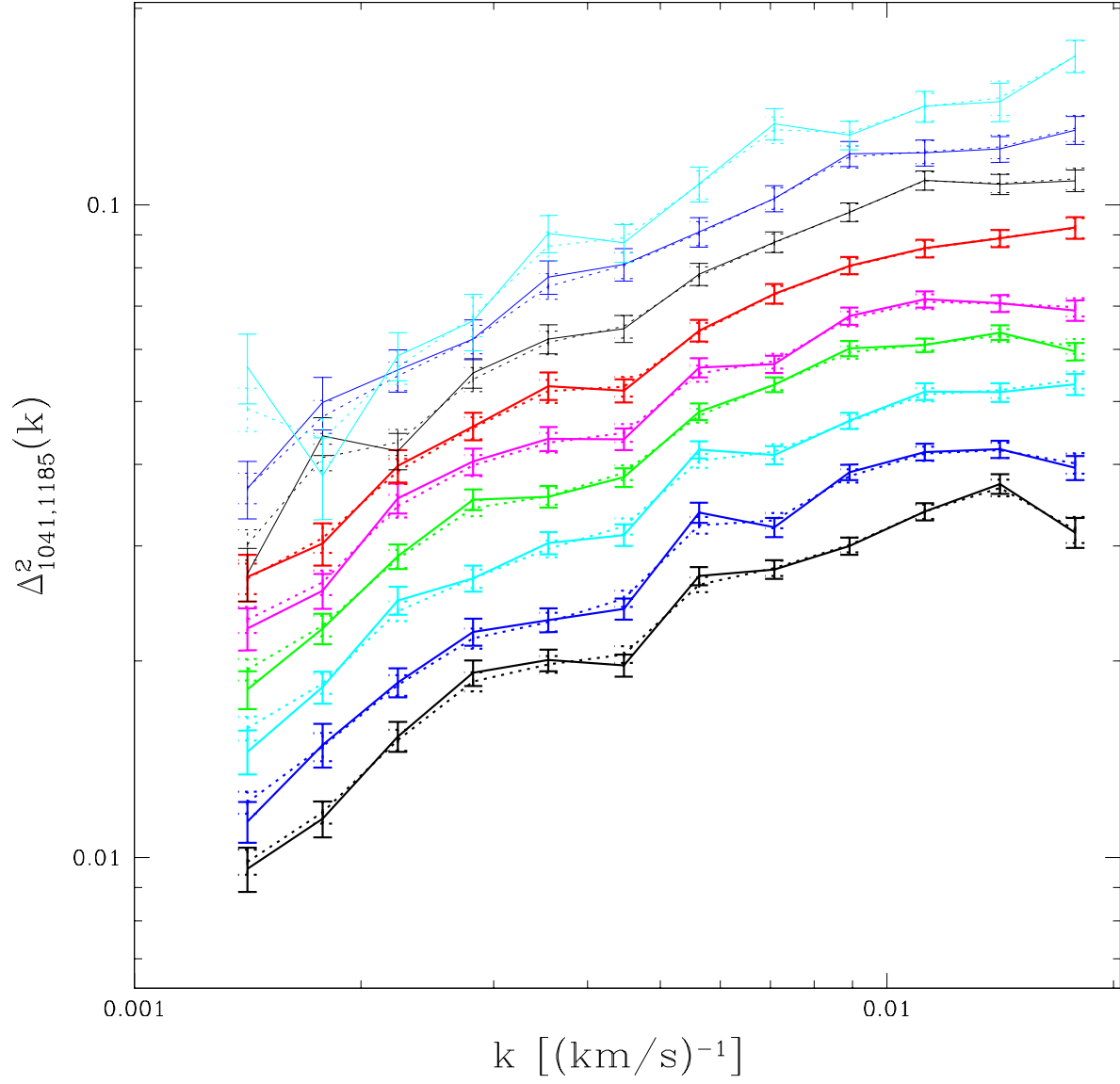


Fig. 24.— Dotted lines connect the power spectrum points before diagonalization of the window matrix. Solid lines show the points after diagonalization.

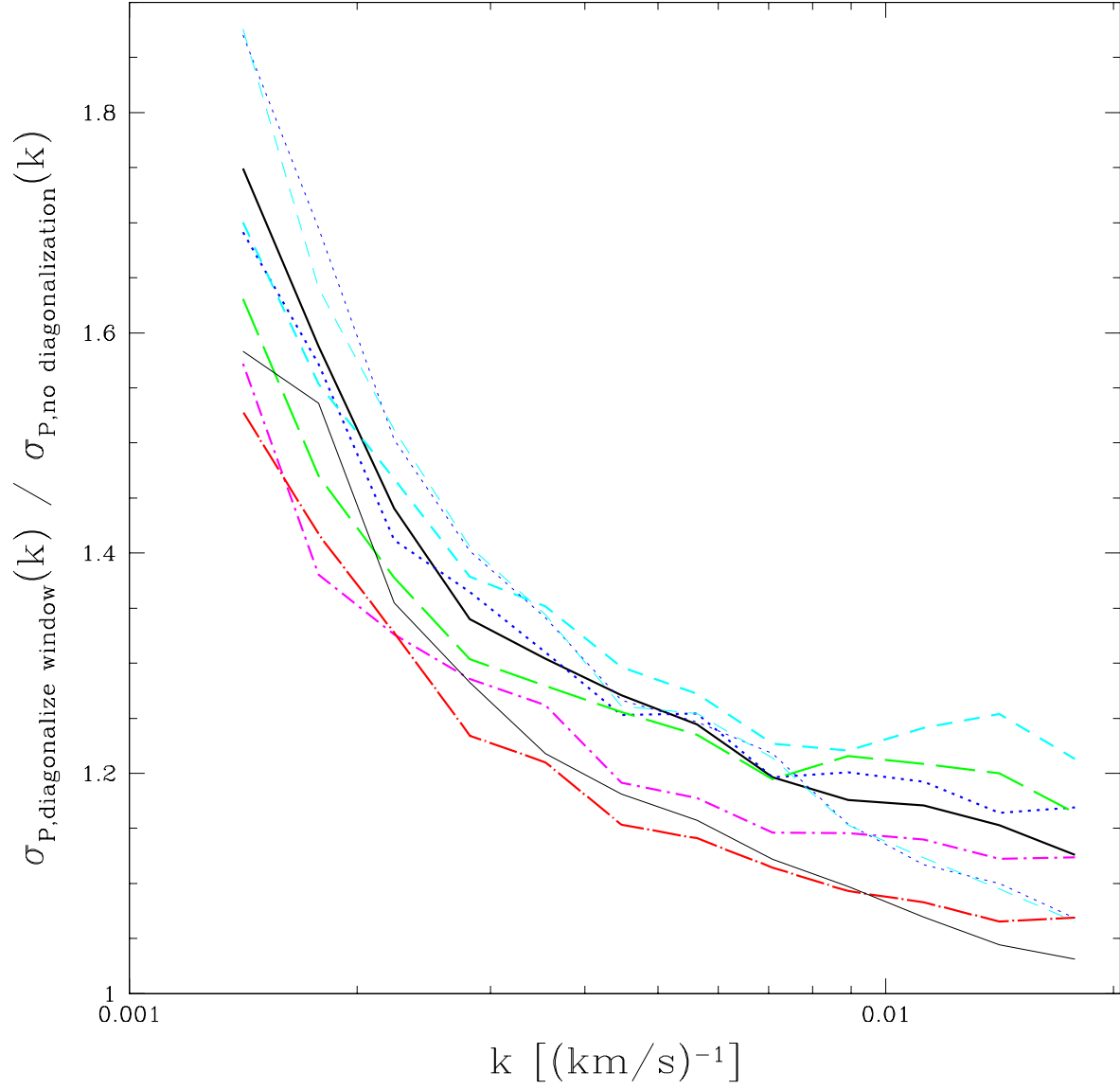


Fig. 25.— Ratio of the diagonal errors on our $P_{1041,1185}(k, z)$ points (connected) after the window matrix diagonalization to before the window matrix diagonalization, with the usual line definitions.

too small (i.e., we are effectively removing 9 data points from the fit so we generally expect a decrease in χ^2 , increase in the error bars, and some corresponding drift in the parameter values). Removing our spectrum-by-spectrum noise estimation makes very little difference. Finally, we note that if we did not correct the noise as discussed in §2.3, the results would change significantly. As an example, we show the fit results in the $\delta \ln \Delta^2 - \delta n_{\text{eff}}$ plane in Figure 26, for our standard case and these noise-related variants. We show the ratio of the power without individual noise estimates for each quasar to our standard case in Figure 27.

Our requirement that the principal components of the error matrix have at least the Gaussian level of variance makes no difference to the fit results, although it improves χ^2 for the fit a little bit. Simply using the Gaussian error matrix instead of bootstrap errors makes no difference to the fit results but increases χ^2 significantly. Ignoring the bootstrap error correlations increases the error on n_{eff} by about 12%, and significantly reduces χ^2 .

We normally use the range $1268 < \lambda_{\text{rest}} < 1380\text{\AA}$ for our background subtraction (i.e., subtract $P_{1268,1380}$). Removing the background subtraction entirely reduces the inferred amplitude by 2σ , and the slope by 1σ , and results in a very large χ^2 (the error on n_{eff} also decreases significantly, but this is mostly because of the change in the best fit values, not because of uncertainty in the background subtraction). (Note that removing the background subtraction, which increases $P_F(k, z)$, decreases the inferred amplitude because the fitted \bar{F} decreases more than enough to offset the increase in $P_F(k, z)$.) Clearly the background cannot be ignored. Using $P_{1409,1523}$ instead produces a somewhat disturbingly large 0.028 (1.1σ) increase in n_{eff} . We expect the longer wavelength range to give a less accurate estimate of the background power, because some metals are missing, but further investigation shows that most of this difference is probably caused by CIV BALs adding power to the $1409 < \lambda_{\text{rest}} < 1523\text{\AA}$ region. As we see in Table 1, removing the adjustment for noise dependence of the background (see Equation 13) brings the two background regions closer together (this is reflected in Figure 18). Adding the 147 BAL quasars identified by our automated algorithm leads to a huge discrepancy (0.094 in n_{eff}) when we use the $P_{1409,1523}$ background, but only when we adjust for noise level (without this the discrepancy for n_{eff} , not shown in the table, is only 0.029). Note that the BAL cut makes essentially no difference to our standard fit using the $P_{1268,1380}$ background. All of these differences are easy to understand: First, BALs are known to be strongest in CIV absorption (Hall et al. (2002)), so it is not surprising that we see the effects of BALs primarily in this wavelength region. Second, both our original by-eye and subsequent automated removal of BALs inevitably identify the features more easily in less noisy data, so the power from BALs naturally shows up when we intentionally use the noisier spectra for the background power. The fact that removing the 147 most obvious BALs has essentially no effect on our basic result gives us confidence that any remaining BAL features in the Ly α forest and $1268 < \lambda_{\text{rest}} < 1380\text{\AA}$ regions are not significant.

To investigate the effect of a systematic uncertainty in the spectral resolution, we include in our fits an overall factor of the form $\exp(\alpha k^2)$ multiplying the power spectrum, where α is a free parameter. In our standard fit we impose an external constraint on α , $\pm(7 \text{ km s}^{-1})^2$ rms. Essentially

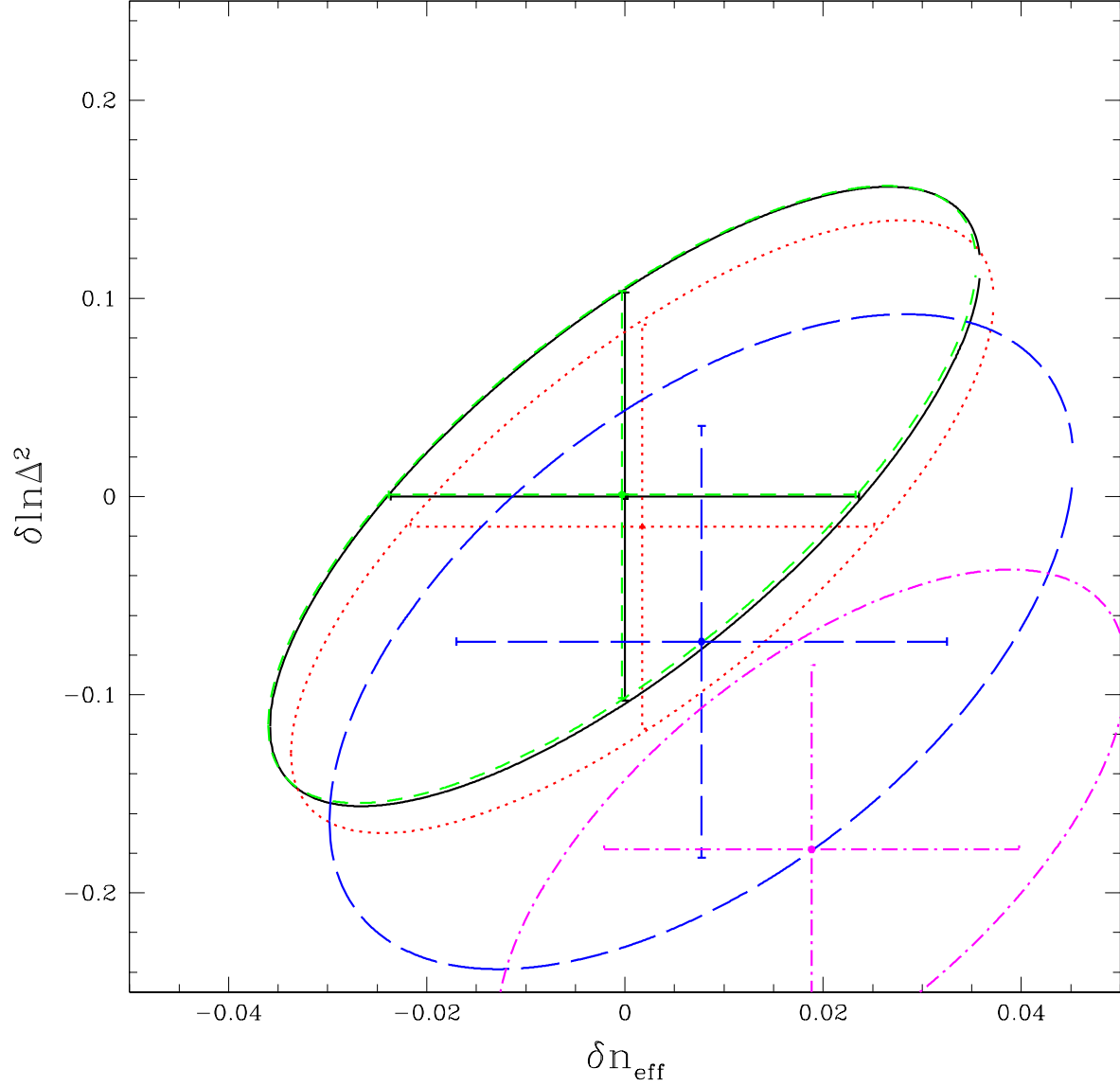


Fig. 26.— Fit results for variant noise treatments. The error bars show the 1σ error on each parameter. The ovals show $\Delta\chi^2 = 2.3$. Standard case (with 5% noise amplitude freedom in each redshift bin): black, solid lines. No individual noise estimate for each quasar: red, dotted lines. Noise amplitude freedom 0.5% (50%): green, short-dashed lines (blue, long-dashed lines). The magenta, dot-dashed line shows the result using the pipeline noise estimates.

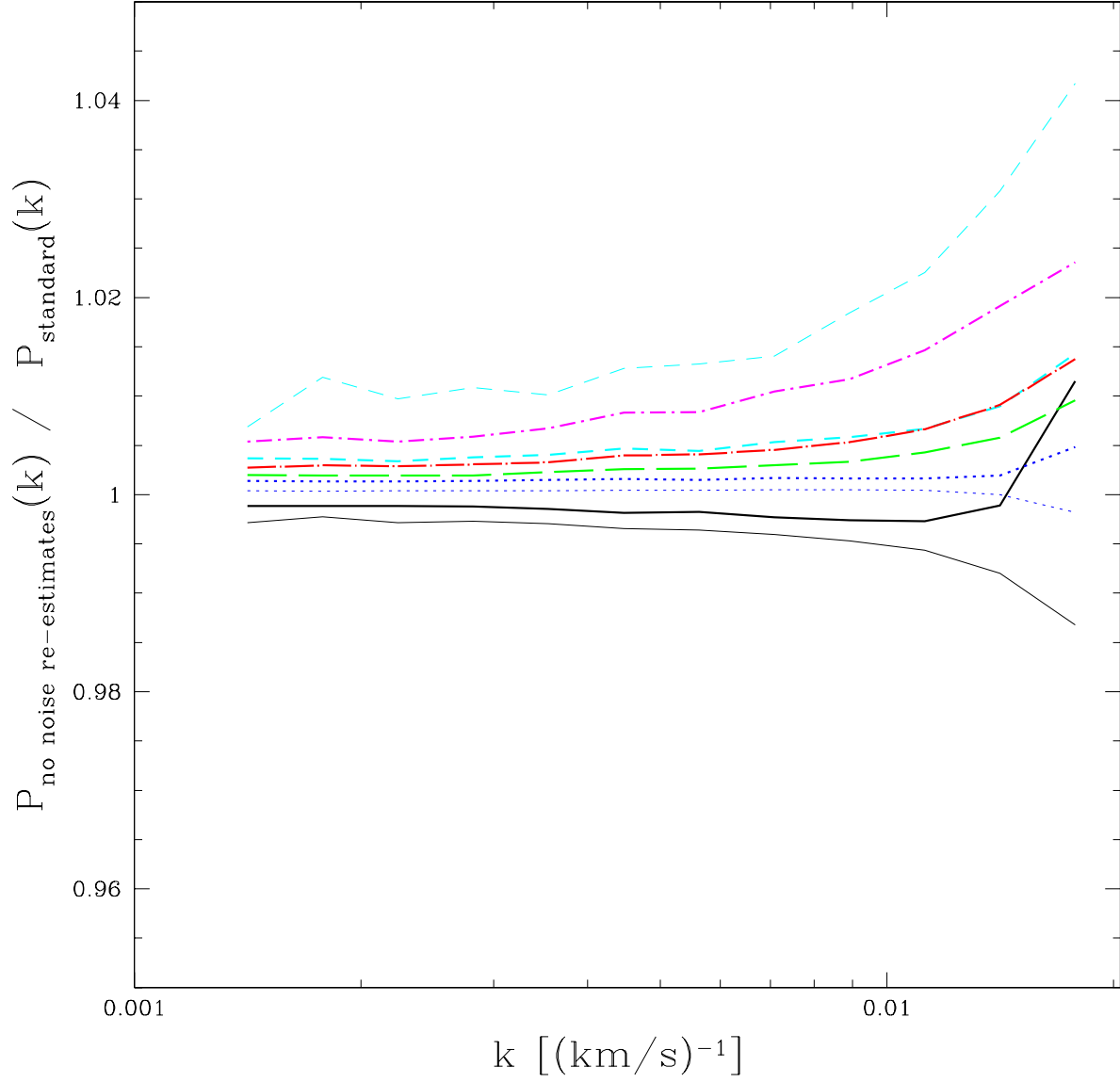


Fig. 27.— Ratio of $P_{1041,1185}$ computed without quasar-by-quasar noise re-estimates (a constant 16% extra noise power was assumed instead) to the standard case.

removing this freedom has no effect on the fit, while leaving α essentially free increases the error on the amplitude by 40%, and increases n_{eff} by $2/3\sigma$ (the change in fitted amplitude is certainly consistent with drift from the increased error). As we show in Figure 5, our standard fit should be conservative.

Simply dividing each chunk of spectrum by its mean instead of also dividing by the mean continuum before estimating the power from the chunk makes little difference to the fit results. Division by the mean continuum actually increases the measured flux power by $\sim 0 - 2\%$, as we show in Fig. 28. We have performed a preliminary PCA analysis to try to model fluctuations around the mean continuum. When we use continua for each quasar composed of 13 PCA eigenvectors, our results change only a little (n_{eff} is reduced by 0.4σ), and χ^2 increases, probably an indication of the unsatisfactory level of noise that we know remains in our estimates. As we see in Figure 29, the modification of adding a large constant to the weight matrix to make our measurement less sensitive to the mean of each chunk has little effect (the effect is larger on larger scales).

The line “no bin-redshift correction” in Table 1 refers to removing the correction for evolution in the power across the width of the redshift bins (see eqn. 8). We see (Fig. 30) that this correction mostly affects the lowest redshift bin (where the low- z edge of the bin is empty of data) and has little effect on the fit (not surprisingly, leaving out this correction increases χ^2).

The line “ignore $F - \sigma_p$ correlation” in Table 1 shows the change in the fitted parameters if we naively use the given noise estimates for weighting without accounting for the fact that there is a correlation between the flux estimate and the noise amplitude estimate for each pixel. Figure 31 shows that the bias is a fairly constant 3-5% increase in the flux power. The difference is not actually caused by the change in weighting used in the power spectrum estimation – instead, the power is biased high because the estimation of the mean that each chunk of spectrum is divided by is biased low because low-flux pixels have smaller noise estimates. Ignoring this effect does not change our fit results. Normally we base our estimation of the amount of the noise that is due to quasar flux on the separate estimates we have from the spectral reduction pipeline for the flux, sky, and read-noise contributions, however these estimates do not add up to the total noise reported by the pipeline. If we rescale the individual numbers to make them consistent with the total (not necessarily the correct thing to do) we see that the fit results are not changed significantly (the line “rescale σ_c ” – we use σ_c to refer to noise computed using the separate flux, sky, and read-noise estimates), although the power does change by as much as 3% (Fig. 32 – this difference would be a bit larger if we did not directly measure and correct for the cross-correlation between the noise amplitude and flux).

Finally, our power spectrum extraction code has a bias (partially related to division by the mean of each chunk of spectrum), that we correct for by dividing the result by Equation 11. Removing this correction decreases the estimated n_{eff} by about $1/2 \sigma$ and increases the amplitude by a similar amount. The combined change is actually quite significant because it is transverse to the degeneracy direction for these parameters.

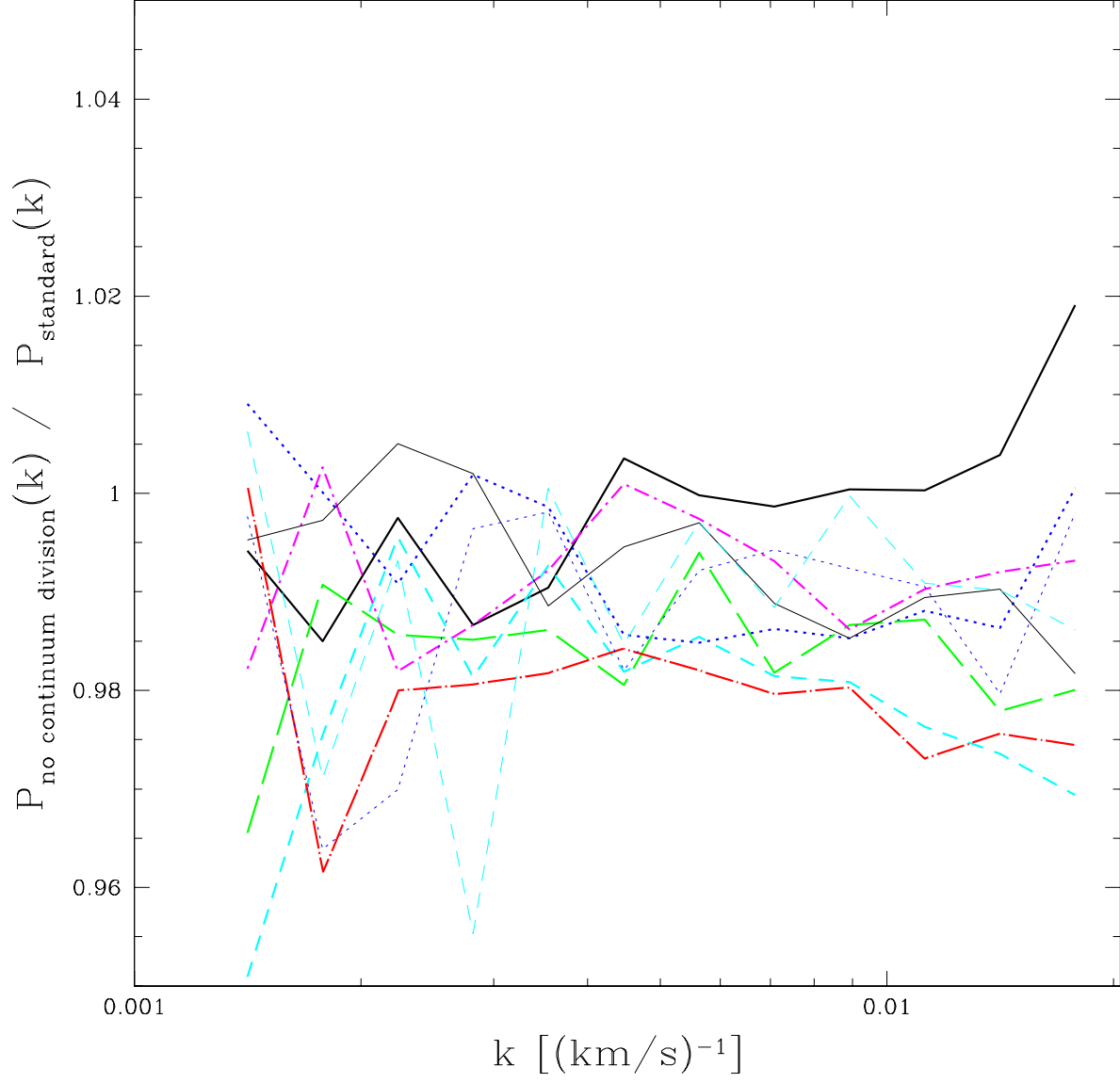


Fig. 28.— Ratio of $P_{1041,1185}$ computed without dividing by the continuum estimate to the standard case.

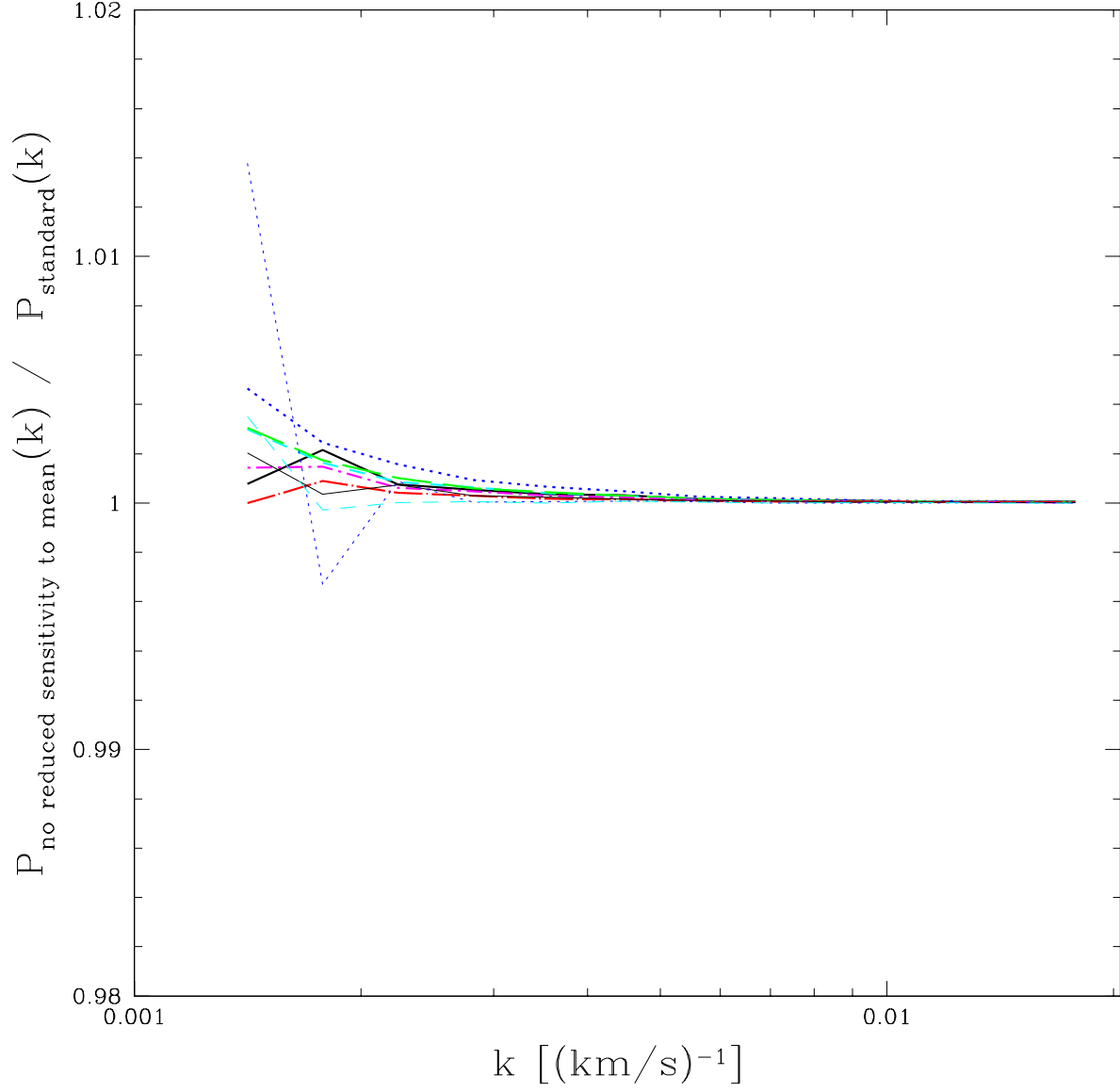


Fig. 29.— Ratio of $P_{1041,1185}$ computed without the large constant added to the weight matrices to make the results less sensitive to the mean of the chunks, relative to the standard case.

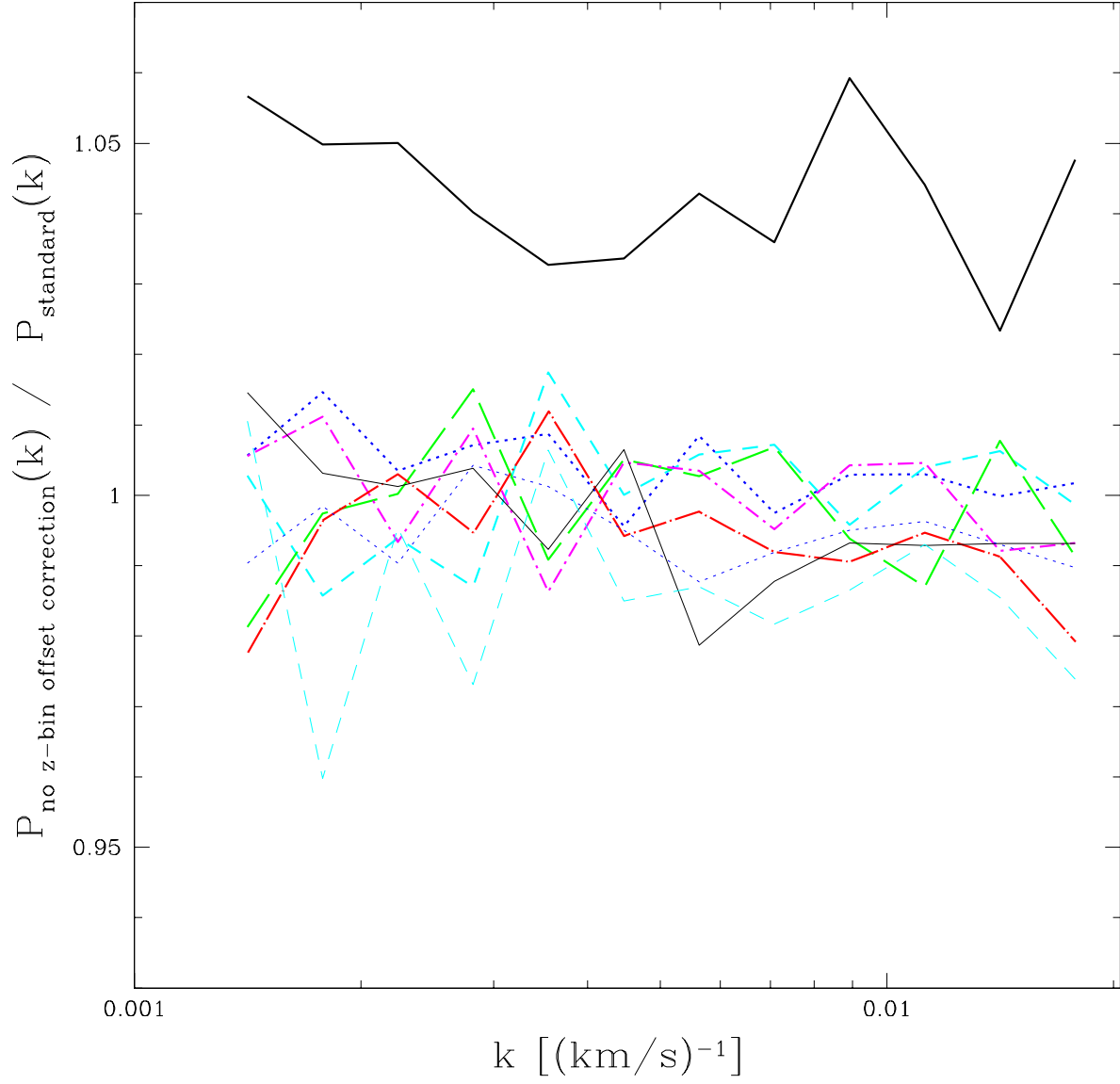


Fig. 30.— Ratio of $P_{1041,1185}$ computed with no correction for the offset between the defined center of each redshift bin and the center of weight of the data to the standard case.

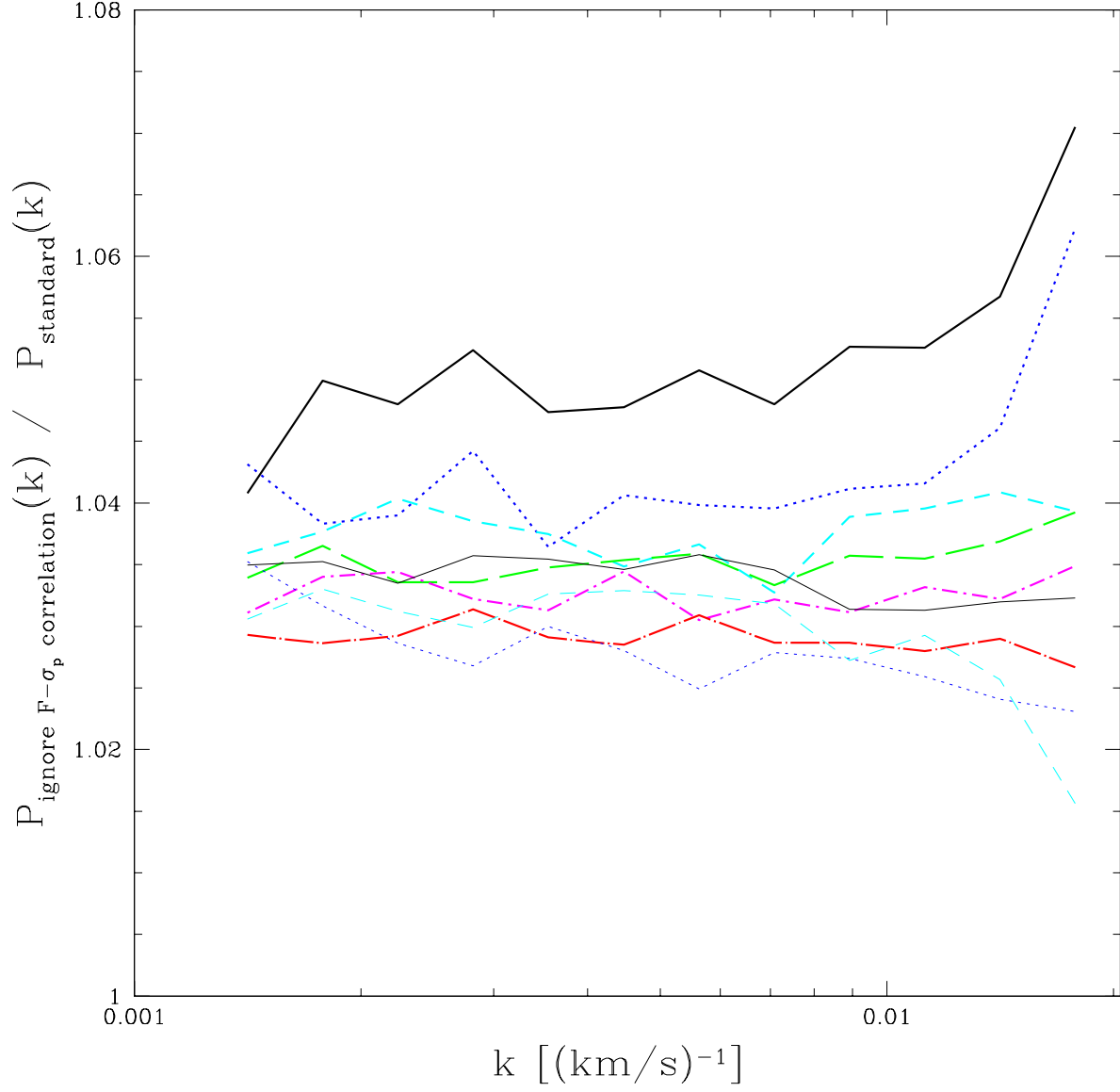


Fig. 31.— Ratio of $P_{1041,1185}$ computed without accounting for the correlation between the noise amplitude and the observed flux to the standard case.

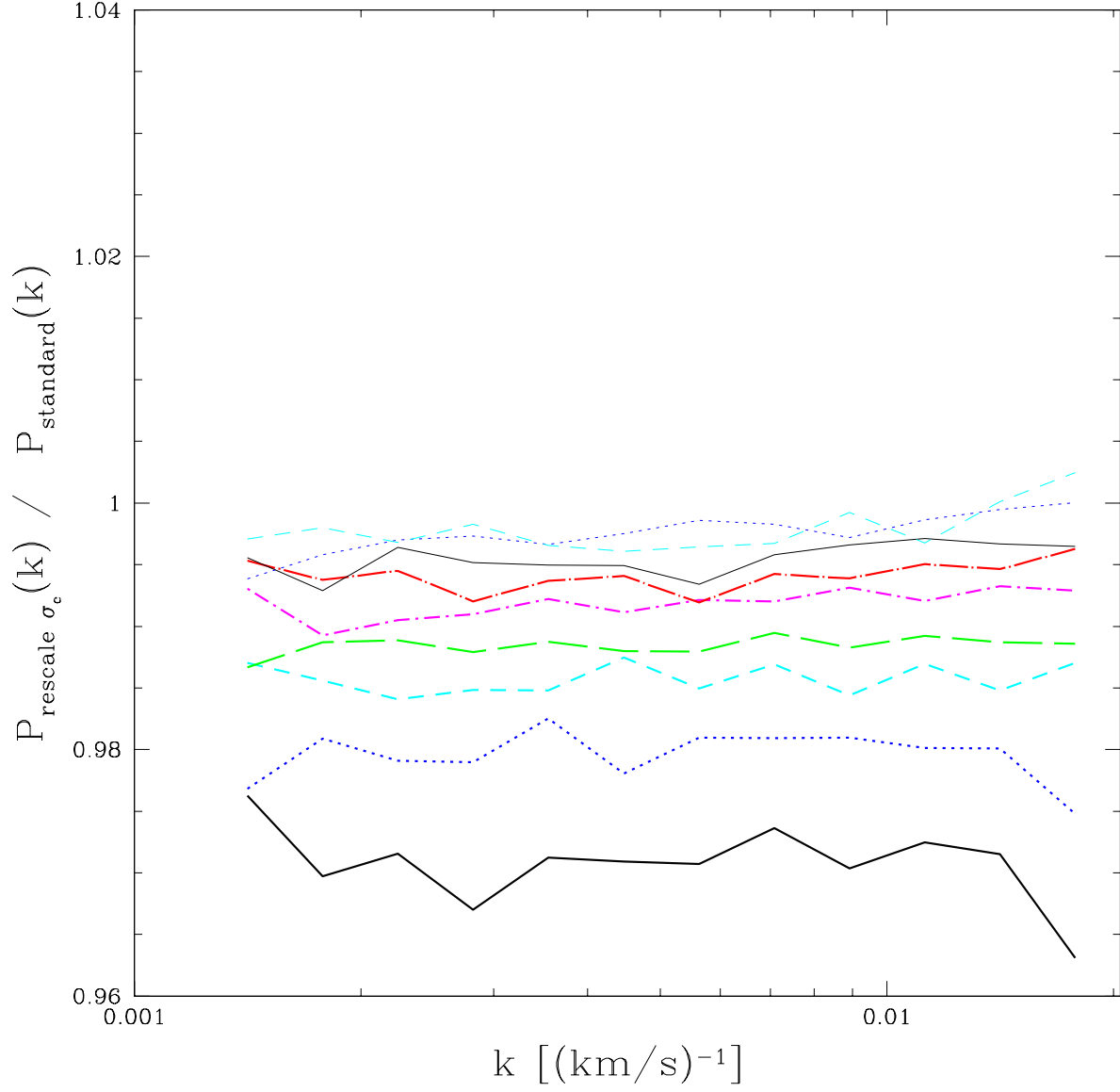


Fig. 32.— Ratio of $P_{1041,1185}$ computed using an alternative estimate of the fraction of the noise that is due to photon counting noise associated with flux from the quasar (see text) to the standard case.

To summarize, most of the effects described above are small relative to the statistical errors on the final estimated parameters. We understand the cases where the difference is significant, and expect that our standard method will be much more accurate than the difference between it and the variants (we show these variants to help the reader better understand our measurement). These tests give us confidence that the final results are very robust to small changes in the analysis pipeline.

How sensitive are these conclusions to our assumptions about the nuisance parameters, \bar{F} , $T_{1.4}$, and γ , i.e., if these constraints improve in the future, will we need to worry more about systematic errors in $P_F(k, z)$? We investigate this by first fixing all the parameters in the fit (including removing the noise amplitude uncertainty, resolution uncertainty, and freedom in the SiIII correction), so the only uncertainty is on $P_F(k, z)$. Table 1 (the “fixed nuisance parameters” line) shows that the error on the amplitude improves dramatically, by a factor of 5. The error on n_{eff} improves by a factor of 2. So in principle the amplitude error can be improved a lot relative to potential systematic errors, and n_{eff} improved as well (see also Mandelbaum et al. 2003). The next line (“optimistic \bar{F} ”), where we assume the HIRES constraint on \bar{F} is improved by a factor of 5, and the SDSS constraint by a factor of 2, shows that Δ^2 is substantially degenerate with \bar{F} (as expected), but n_{eff} is less degenerate. Improving the constraints on $T_{1.4}$ and γ by factors of 5, in addition to the improved constraints on \bar{F} , leads to little further improvement. Finally, for comparison, we tried simply reducing the errors on $P_F(k, z)$ by a factor of $\sqrt{3}$, and found that the error on n_{eff} decreases by almost the same factor (1.6), but the error on Δ^2 decreases less (a factor of 1.2). SDSS will collect a factor of ~ 3 more data than we have in the present sample. We conclude that the error on n_{eff} can easily be reduced by simply gathering more data, while improvements on Δ^2 can be made by improving the errors on \bar{F} .

Table 1. Effect of uncertainties or modifications of the $P_F(k, z)$ measurement procedure on the inferred linear power spectrum

Variant ^a	$\delta \ln \Delta^2$	$\sigma_{\ln \Delta^2}$	δn_{eff}	$\sigma_{n_{\text{eff}}}$	$\Delta \chi^2$ ^b	χ^2 ^c
Standard fit	0	0.10	0	0.024	0	129.7
No window diagonalization	-0.06	0.10	-0.024	0.021	1.4	129.5
No SiIII correction	0.04	0.10	0.017	0.021	0.7	193.7
z -independent SiIII	0.00	0.10	0.0	0.024	0.0	130.9
variable width SiIII	-0.02	0.11	-0.006	0.025	0.1	129.0
Include SiIII-SiIII term	-0.02	0.10	-0.003	0.023	0.0	132.1
$\sigma_{\text{noise power}} = 0.5\%$	0.00	0.10	-0.000	0.024	0.0	130.3
$\sigma_{\text{noise power}} = 50\%$	-0.08	0.12	0.008	0.025	1.1	123.8
No individual noise re-estimation	-0.02	0.10	0.002	0.023	0.1	128.1
Believe pipeline noise	-0.20	0.11	0.019	0.021	9.6	129.9
No Gaussian floor on errors	0.01	0.10	0.002	0.024	0.0	133.1
Gaussian errors	0.02	0.10	0.001	0.023	0.1	151.7
Ignore error correlations	0.04	0.11	0.002	0.027	0.2	117.2
No background subtraction	-0.20	0.10	-0.022	0.019	3.6	169.6
Background 1409-1523Å	0.05	0.10	0.028	0.025	1.5	133.2
No background noise matching	-0.07	0.10	-0.004	0.021	0.7	142.3
Previous, but use 1409-1523Å	-0.06	0.10	0.008	0.022	1.7	143.1
No automated BAL cut	0.01	0.10	0.003	0.023	0.0	127.0
Previous, but use 1409-1523Å	0.14	0.10	0.094	0.025	16.0	156.6
$(70 \text{ km s}^{-1})^2$ resolution error	-0.11	0.14	0.015	0.024	1.4	126.4
$(0.7 \text{ km s}^{-1})^2$ resolution error	0.02	0.10	-0.003	0.024	0.3	130.4
No continuum division	0.02	0.10	0.002	0.024	0.1	132.2
PCA continuum division	0.02	0.10	-0.010	0.024	0.8	139.1
No reduced sensitivity to mean	-0.00	0.10	-0.002	0.024	0.0	130.5
No bin-redshift correction	-0.02	0.10	-0.006	0.023	0.1	137.2
Ignore $F - \sigma_p$ correlation	0.00	0.10	0.002	0.024	0.0	128.7
rescale σ_c	0.00	0.10	0.002	0.024	0.0	129.0
No code bias correction	0.06	0.10	-0.013	0.024	3.0	132.3
8000 bootstrap sets	0.00	0.10	0.000	0.024	0.0	128.5
fixed nuisance parameters	0.083	0.021	-0.025	0.012	—	—
optimistic \bar{F}	0.068	0.062	0.009	0.019	—	—
optimistic $T_{1.4}, \gamma$	0.002	0.082	-0.014	0.021	—	—
optimistic $\bar{F}, T_{1.4}, \gamma$	0.002	0.051	-0.016	0.018	—	—

Table 1—Continued

Variant ^a	$\delta \ln \Delta^2$	$\sigma_{\ln \Delta^2}$	δn_{eff}	$\sigma_{n_{\text{eff}}}$	$\Delta\chi^2$ ^b	χ^2 ^c
$P_F(k, z)$ errors divided by $\sqrt{3}$	-0.051	0.081	-0.000	0.015	—	—

Note. — $z_p = 2.6$, $k_p = 0.009 \text{ (km s}^{-1}\text{)}^{-1}$.

^aThe meaning of each variant is explained in §4.3.

^b $\Delta\chi^2$ of the fitted parameters relative to the standard parameters, using the errors from the variant fit.

^c χ^2 for the fit (essentially unrelated to $\Delta\chi^2$).

4.4. Subsamples of the Data

Another way to test for systematic errors is to search for internal discrepancies between the different subsamples of the same data. Of course, there are only a finite number of possible subsamples we can try, so this test cannot be fully exhaustive. In addition, with many such tests performed one must worry that some will give an apparently statistically significant deviation just by random chance. Table 2 shows results of splitting the data into roughly equal weight subsamples, defined by various properties of the spectra that, at least at first glance, should not be correlated with the measured power. In practice, we rank the spectra by the property of interest and split the sample into halves by requiring that the Gaussian errors on the $k = 0.007 \text{ (km s}^{-1}\text{)}^{-1}$ point are equal for the two halves (the bootstrap errors will not be precisely equal). We list the probability of obtaining χ^2 greater than the value computed by differencing the power spectra (these differences include the different background subtraction computed using eqn. 13 for different noise levels). We also list the fitting parameter results for each subsample, and give the probability for obtaining the observed level of difference between the fits. Because these subsamples are basically independent, deviations within the error bars are expected and are not an indication of systematic errors. We describe these subsample splits below.

The power we measure should be independent of the rest frame region of the quasar continuum in which it is measured. Figure 33 shows $P_{1041,1113}(k, z)$ and $P_{1113,1185}(k, z)$ to test this expectation. The results look pretty similar, but to compare them quantitatively, we compute $\chi^2 = (\mathbf{P}_< - \mathbf{P}_>)^T (\mathbf{C}_< + \mathbf{C}_>)^{-1} (\mathbf{P}_< - \mathbf{P}_>)$, finding $\chi^2 = 111.0$ for 108 points. The agreement appears perfect. To compare the two in a different way, we perform separate fits of the linear mass power spectrum parameters Δ^2 and n_{eff} to $P_F(k, z)$ computed from $P_{1041,1113}(k, z)$ and $P_{1113,1185}(k, z)$. The results, given in the first line of Table 2, are consistent within the expected errors. This test provides some evidence that power from continuum fluctuations is not an important contribution to the total, beyond what we would expect from looking at the red side of the Ly α emission line. It is possible that the two halves of the forest could have significant extra continuum power, but if they do it has to be the same in each half.

We compute the weighted mean of the rms noise for each chunk of spectrum as we use it to estimate the power spectrum. A split based simply on this noise level, illustrated in Figure 34, produces a small but unambiguously significant discrepancy in the raw measurement of $P_{1041,1185}$, $\chi^2 = 185$, though the fit parameters agree within their errors (Table 2, line 3). This discrepancy in power is the motivation for, and is largely removed by, our noise-dependent background subtraction procedure defined by equation (13). Figure 35 shows the power $P_{1268,1380}$ that is used for background subtraction, again subsampled by noise level. There is a clear difference in power, which is not isolated to a few wavenumbers or redshifts. Once the $P_{1268,1380}$ power is subtracted according to equation (13), we obtain P_F estimates and corresponding fit parameters from the high and low noise subsamples that agree within the errors (Table 2, line 2). Since the fit parameters agree even without noise-dependent background subtraction, it appears that the discrepancy in raw $P_{1041,1185}$ power does not mimic a change in cosmological parameters, and our ultimate conclusions would

Table 2. Comparison between subsamples of the data

Split	$P_{>\chi^2}^a$ (points)	$\delta \ln \Delta_{<}^2{}^b$	$\delta n_{\text{eff},<}$	$\delta \ln \Delta_{>}^2$	$\delta n_{\text{eff},>}$	$P_{>\chi^2}$ (fit)
λ_{rest}	40%	-0.03 ± 0.12	-0.027 ± 0.031	0.04 ± 0.11	0.018 ± 0.029	51%
noise	10%	-0.01 ± 0.11	0.020 ± 0.028	-0.01 ± 0.12	-0.001 ± 0.030	76%
noise (raw) ^c	0.0006%	-0.05 ± 0.11	0.027 ± 0.029	-0.19 ± 0.13	-0.010 ± 0.026	61%
sky	5.9%	-0.02 ± 0.11	-0.001 ± 0.028	0.02 ± 0.12	0.014 ± 0.030	93%
$\sigma_w - \sigma_c$	34%	0.02 ± 0.11	0.029 ± 0.030	-0.01 ± 0.12	-0.011 ± 0.030	49%
read noise	94%	0.08 ± 0.11	0.020 ± 0.030	-0.05 ± 0.12	-0.015 ± 0.028	68%
cont. χ^2/ν	33%	-0.04 ± 0.11	0.017 ± 0.028	0.06 ± 0.12	0.001 ± 0.030	40%
resolution	73%	0.08 ± 0.11	0.036 ± 0.031	-0.08 ± 0.12	-0.025 ± 0.028	32%
flexure	14%	0.10 ± 0.11	0.040 ± 0.030	-0.11 ± 0.11	-0.033 ± 0.026	19%
alignment	29%	0.09 ± 0.11	0.031 ± 0.030	-0.09 ± 0.11	-0.031 ± 0.027	29%
exp. χ^2/ν	65%	0.01 ± 0.11	-0.015 ± 0.029	-0.00 ± 0.12	0.010 ± 0.030	63%
error on mean	56%	-0.07 ± 0.10	-0.014 ± 0.030	0.07 ± 0.11	0.018 ± 0.031	67%
error on A_q	40%	0.06 ± 0.09	0.020 ± 0.027	-0.01 ± 0.12	-0.014 ± 0.030	68%

^aProbabilities may not be fully reliable because we have not demonstrated that χ^2 is properly distributed.

^bThe subsample fit results cannot be combined to produce the result of the fit to the full data set because the underlying nuisance parameters were not required to be the same.

^cThe “noise (raw)” line shows the comparison without accounting for the noise dependence of the background.

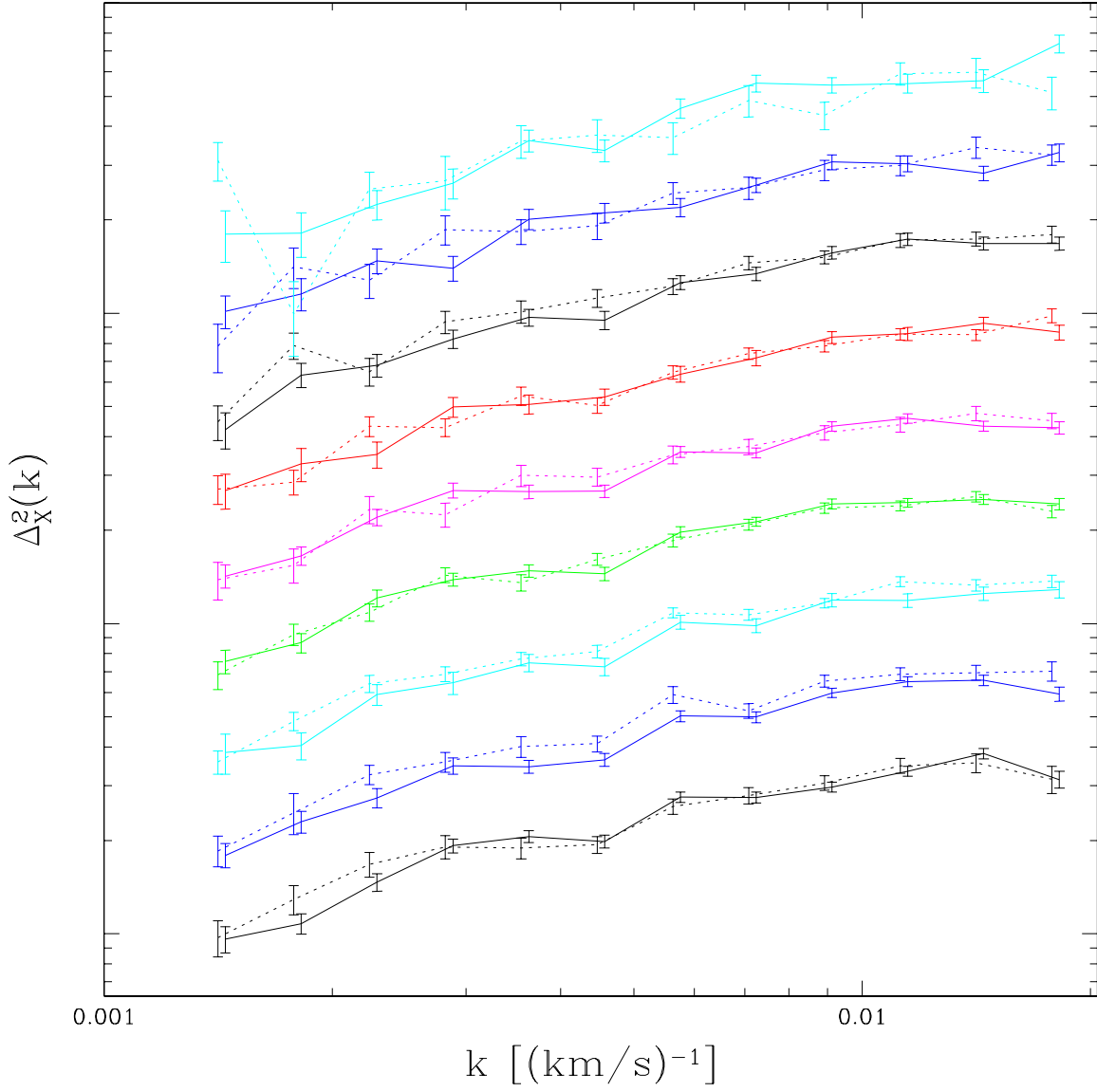


Fig. 33.— Comparison of $P_{1041,1113}(k, z)$ (connected by dotted lines) to $P_{1113,1185}(k, z)$ (connected by solid lines, and shifted slightly to the right). The different redshifts have been shifted vertically by arbitrary amounts on this logarithmic plot (z increases from bottom to top).

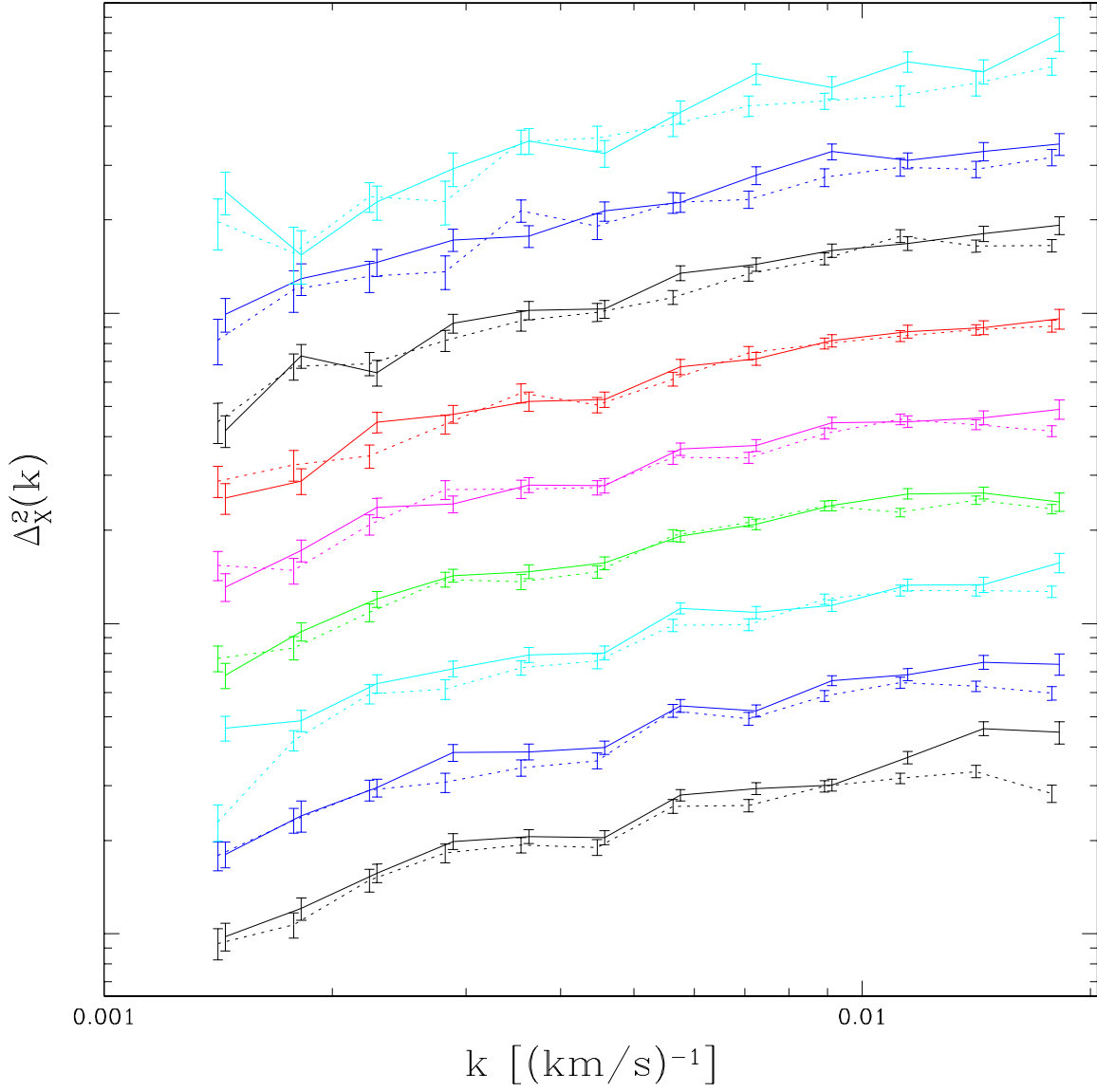


Fig. 34.— $P_{1041,1185}$ split by noise in the Ly α forest region. The dotted line connects the low noise results, while the high noise results are offset slightly to the right. The results at different redshifts have been shifted vertically by arbitrary amounts (z increases from bottom to top).

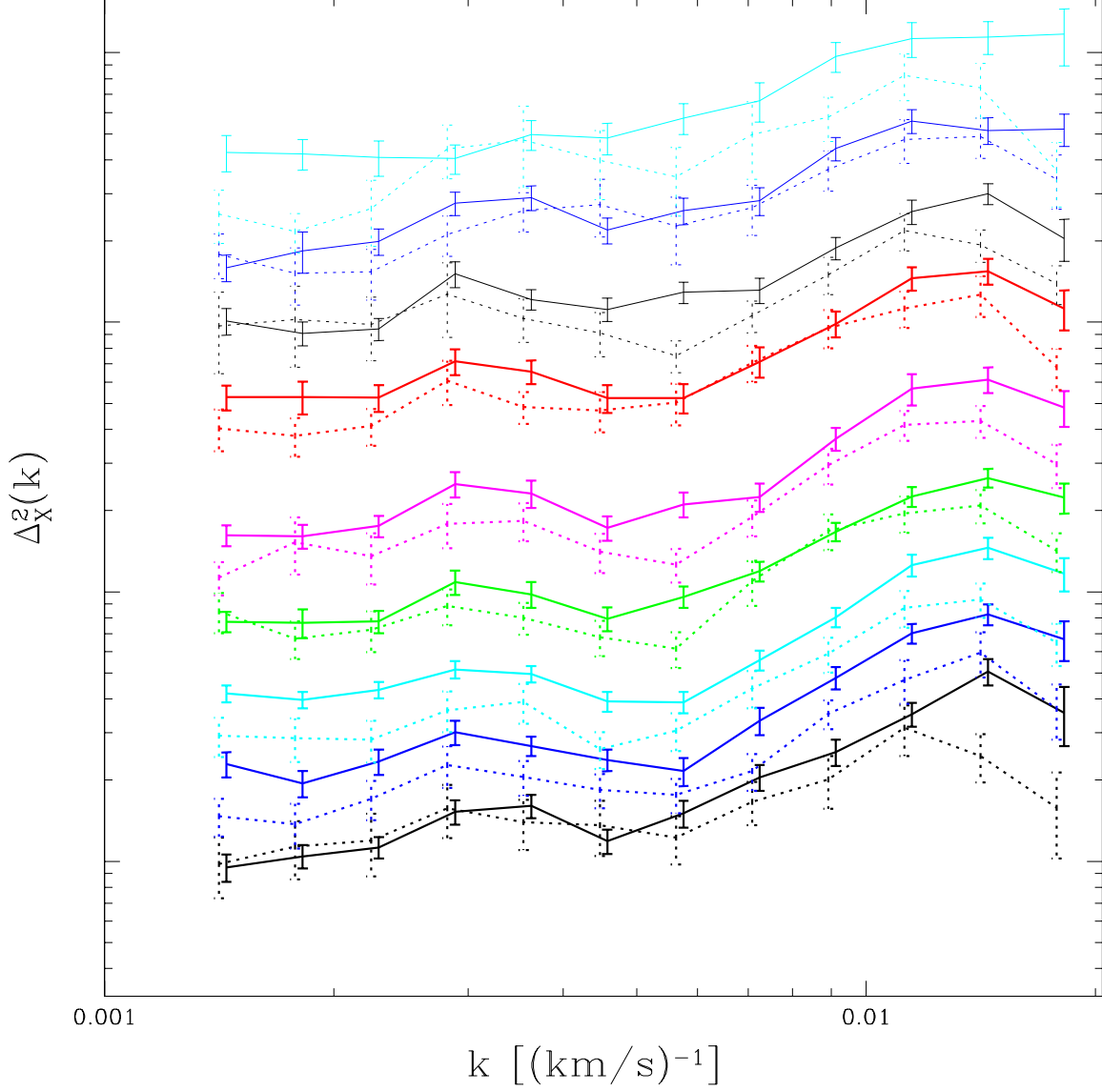


Fig. 35.— $P_{1268,1380}$ split by noise in the same region. The dotted line connects the low noise results, while the high noise results are offset slightly to the right. The results at different redshifts have been shifted vertically by arbitrary amounts.

therefore not change even if we did not implement it. Nonetheless, the origin of the difference remains somewhat mysterious, since we went to great effort to estimate noise correctly.

Two more splits that yield discrepant $P_{1041,1185}$ but show no sign of trouble after the noise-dependent background subtraction are based on the ratio of the mean sky flux to the mean quasar flux and on $\sigma_w - \sigma_c$, the difference between the pipeline estimate of the noise and the sum of our estimates of the quasar flux, sky flux, and read-noise components of the noise. We are not sure what $\sigma_w - \sigma_c$ means, since we do not understand the source of noise misestimation in the standard pipeline. Even without noise dependent background subtraction, the fit results did not differ significantly in either of these cases. They are almost surely symptoms of the same noise-related problem discussed above.

The split based on read-noise in the spectra shows good agreement between the $P_F(k, z)$ measurements, even without noise dependent background subtraction as does a split based on how well the mean continuum matches the quasar spectrum outside the Ly α forest, quantified by computing χ^2/ν for the difference between the continuum and spectrum (“cont. χ^2/ν ” in Table 2). Several other splits that show little or no sign of trouble are based on: the mean value of χ^2/ν computed for each pixel when combining exposures (this was the comparison that motivated our spectrum-by-spectrum noise re-estimation), the mean resolution, the movement of the spectrum relative to the detector pixel grid during the observation (“flexure”), the alignment of the pixels in the different exposures for the same spectrum (closely related to flexure), the error on the overall normalization of the spectrum, A_q (see eqn. 2, this error is set by a combination of the noise level outside the forest and the length of spectrum observed outside the forest), and the error on the means computed for the forest chunks (differences at fixed z are related to the length of the chunk and the noise in the forest).

Overall, the agreement between our subsamples is excellent, both for the $P_F(k, z)$ results and the fit results. In some cases this agreement relies on the noise-dependent background subtraction, which we would like to understand better (in no case does the fit agreement rely on this).

4.5. Continuum Power

The power in the mean continuum, for the 4 different rest frame regions identified in Figure 2, is shown in Figures 36a-d, relative to the Ly α forest power (the mean continuum power was measured by replacing the quasar flux in each pixel by the mean continuum level at that pixel). The mean continuum is very well behaved over the k range that we use ($0.0013 - 0.02 \text{ (km s}^{-1}\text{)}^{-1}$), but its fluctuations quickly become significant at $k \lesssim 0.001 \text{ (km s}^{-1}\text{)}^{-1}$. What little power the mean continuum shows in our chosen k range should be removed when we divide the spectra by the continuum; it is only fluctuations around the mean that matter.

We summarize our strong, but maybe not airtight, argument for believing that continuum fluctuations are not corrupting our measurement as follows:

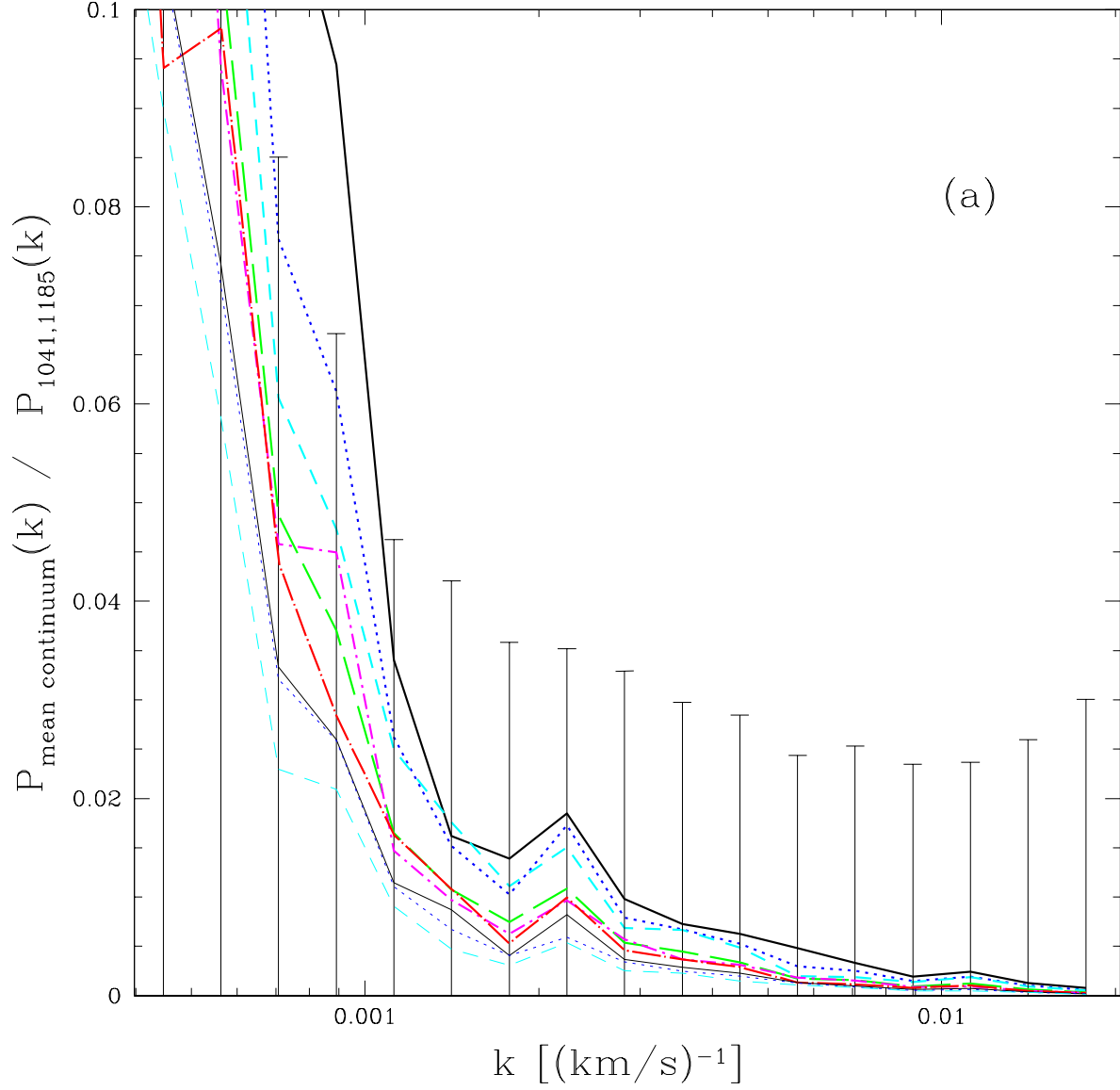
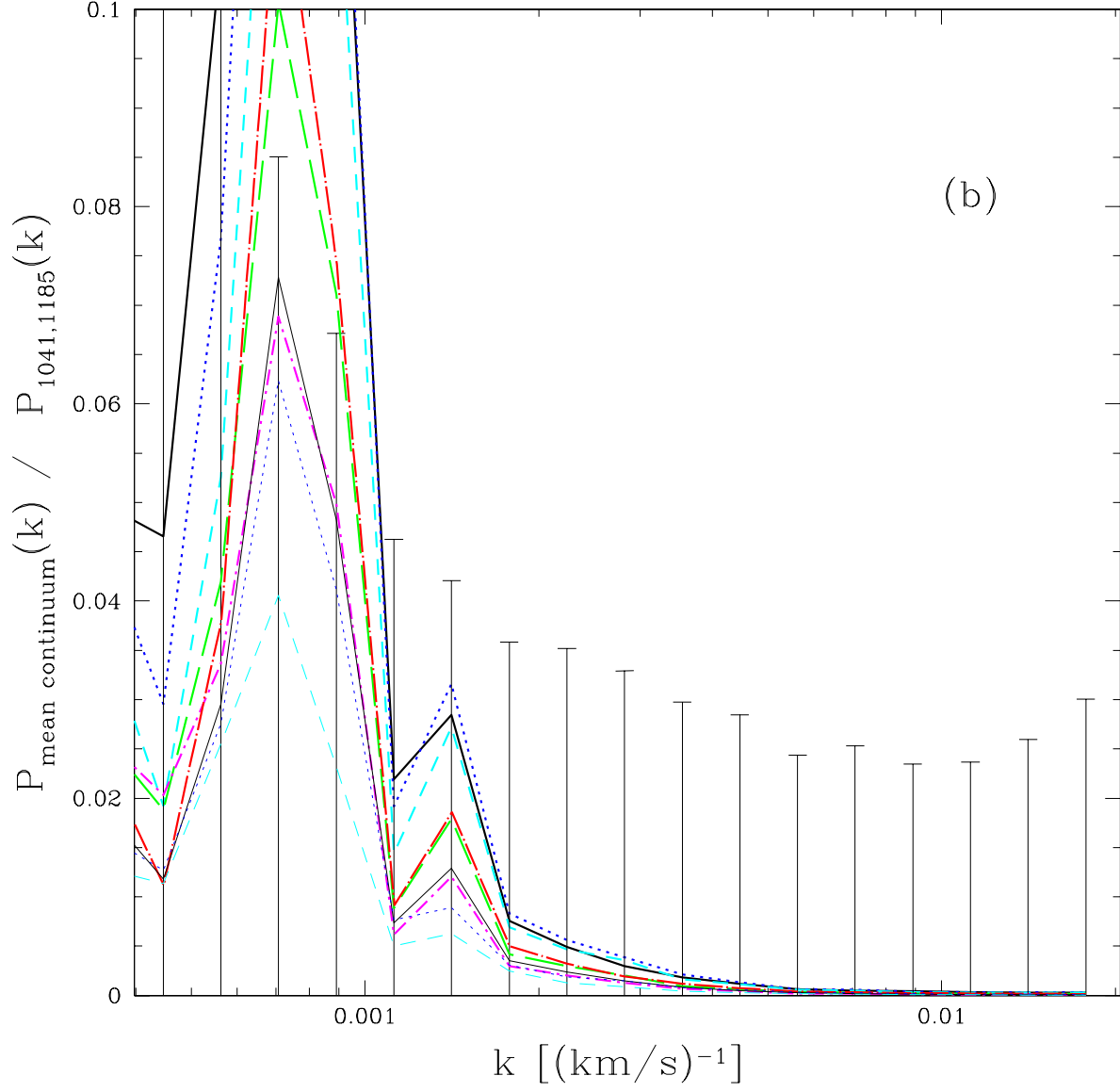
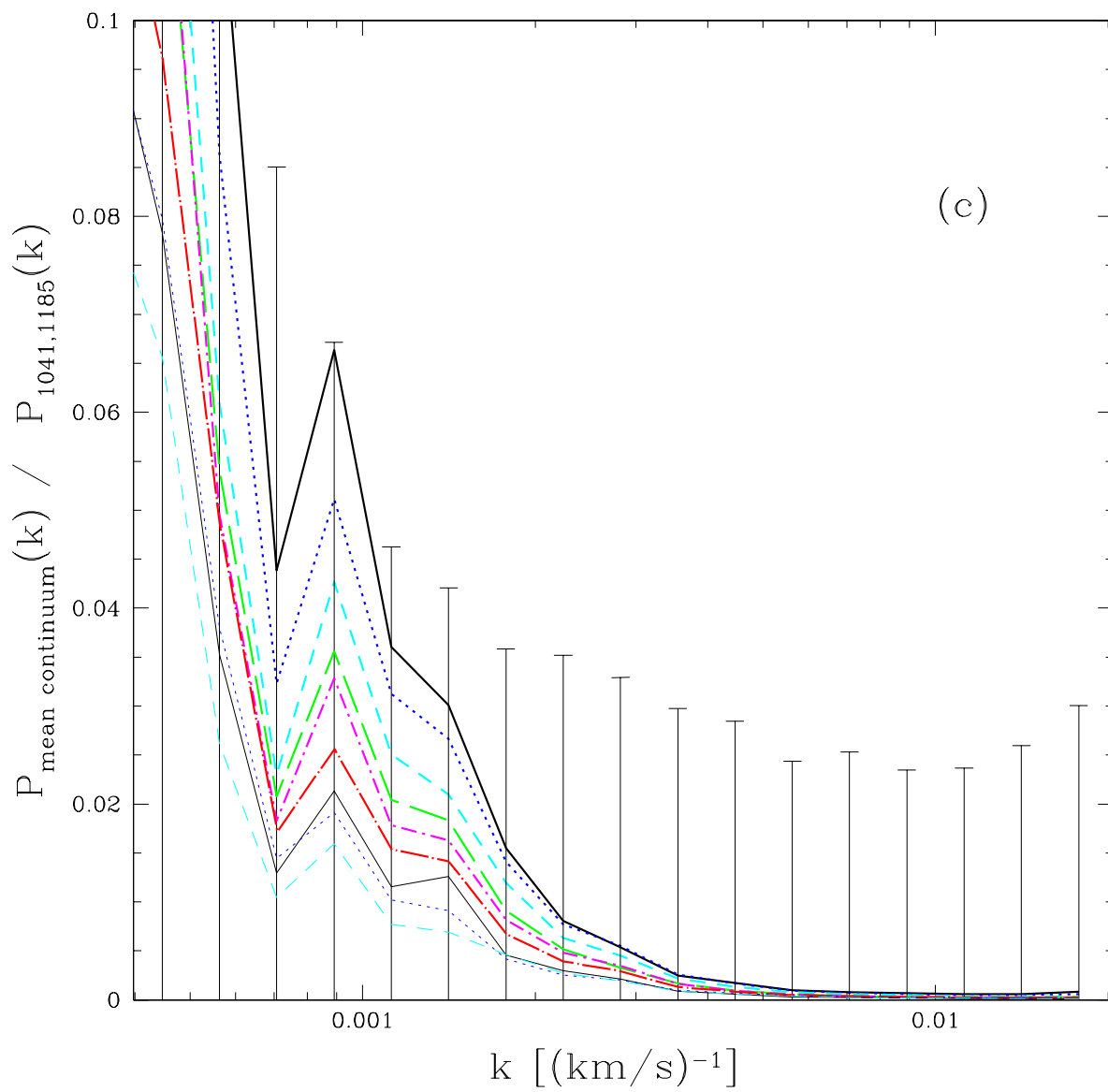
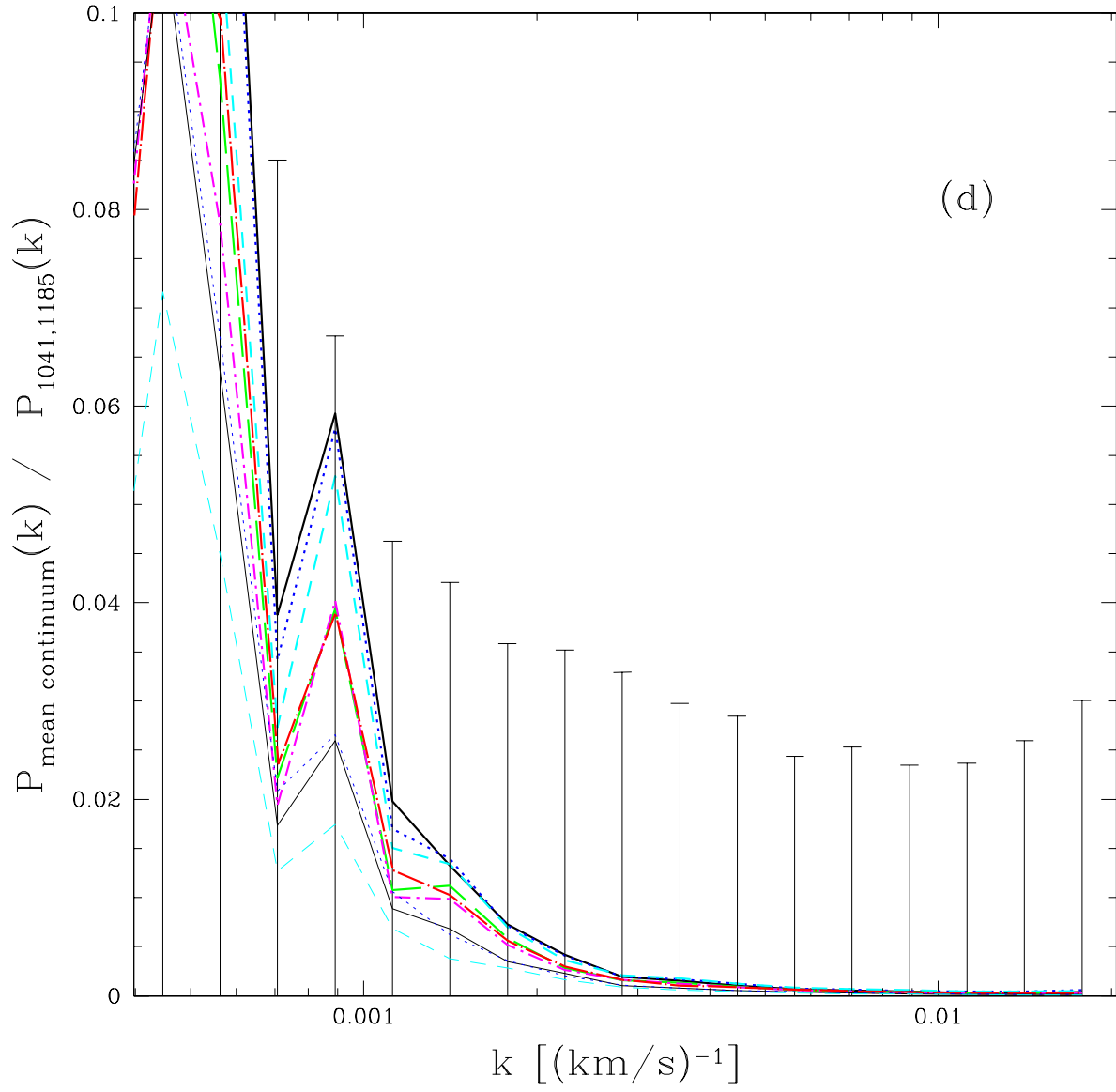


Fig. 36.— Power in the mean continuum relative to the Ly α forest power, for various rest wavelength intervals: (a) $1041 < \lambda_{\text{rest}} < 1185\text{\AA}$, (b) $1268 < \lambda_{\text{rest}} < 1380\text{\AA}$, (c) $1409 < \lambda_{\text{rest}} < 1523\text{\AA}$, (d) $1558 < \lambda_{\text{rest}} < 1774\text{\AA}$. The error bars show the fractional error on $P_{1041,1185}$ (without diagonalizing the window matrix because the diagonalization works poorly at the lowest k s that we show).







- The power in the mean continuum is small.
- The results for the $P_F(k, z)$ measurement in two halves of the forest region, $P_{1041,1113}$ and $P_{1113,1185}$, agree.
- The power in the background regions, $P_{1268,1380}$ and $P_{1409,1523}$, agree at the level we care about in the low-noise data, as long as BALs (which mostly affect the latter region) are removed.
- Division by preliminary 13 eigenvector PCA estimates of the continua (i.e., including fluctuations around the mean) does not change the results.

To be quantitatively important despite these arguments, the power in quasar-to-quasar continuum fluctuations in the forest must be substantially larger than the power in the mean continuum itself, the continuum fluctuations in the forest must be substantially different from those in the background regions (despite those regions being similar to each other and the two halves of the forest being similar to each other), and our PCA analysis must be substantially flawed. Further study is warranted, but a big effect seems unlikely.

4.6. Comparison with Past Measurements

There are three $P_F(k, z)$ measurements already in the literature, McDonald et al. (2000), Croft et al. (2002), and Kim et al. (2004), all using at least some high resolution data. Each uses its own set of redshift bins, so to compare we need a way to interpolate our results to these redshifts. We do this by performing our standard cosmological fit to all of the data (at first – later we will remove some of the past results). This gives us a set of best fit model parameters that can be used to compute the power at any k and z . Within the range of k where we have SDSS measurements, the fit is always dominated by the SDSS points. The fitted curves always match the SDSS results to much better than the size of the errors on the past results, meaning that, for the the purpose of comparison to the past results, the curves are simply a faithful interpolation between the SDSS points. At $k > 0.02(\text{km s}^{-1})^{-1}$, the fit is effectively a weighted average of the past results, although the constraint that it must match SDSS at lower k has some influence (our simulation predictions do not allow for sharp features in $P_F(k, z)$.)

We first perform a fit to all the data with $k < 0.05(\text{km s}^{-1})^{-1}$ and $z > 2.1$, finding an atrociously bad $\chi^2 = 392$ for ~ 238 dof. Removing McDonald et al. (2000), reduces χ^2 by 53.4 (for 39 data points), removing Croft et al. (2002) reduces χ^2 by 85.2 (for 65 points), and removing Kim et al. (2004) reduces χ^2 by 123.3 (28 points). Clearly there is gross disagreement between Kim et al. (2004) and the other results. Figure 37 shows the Kim et al. (2004) points at $z = 2.18$ and $z = 2.58$ (from their Table 5) along with the fit prediction for them. Note that we include SiIII contamination in the model as described in §4.2, so the model curves are not perfectly smooth. We see large discrepancies, as we expect from the bad χ^2 . The point at $k = 0.0012(\text{km s}^{-1})^{-1}$,

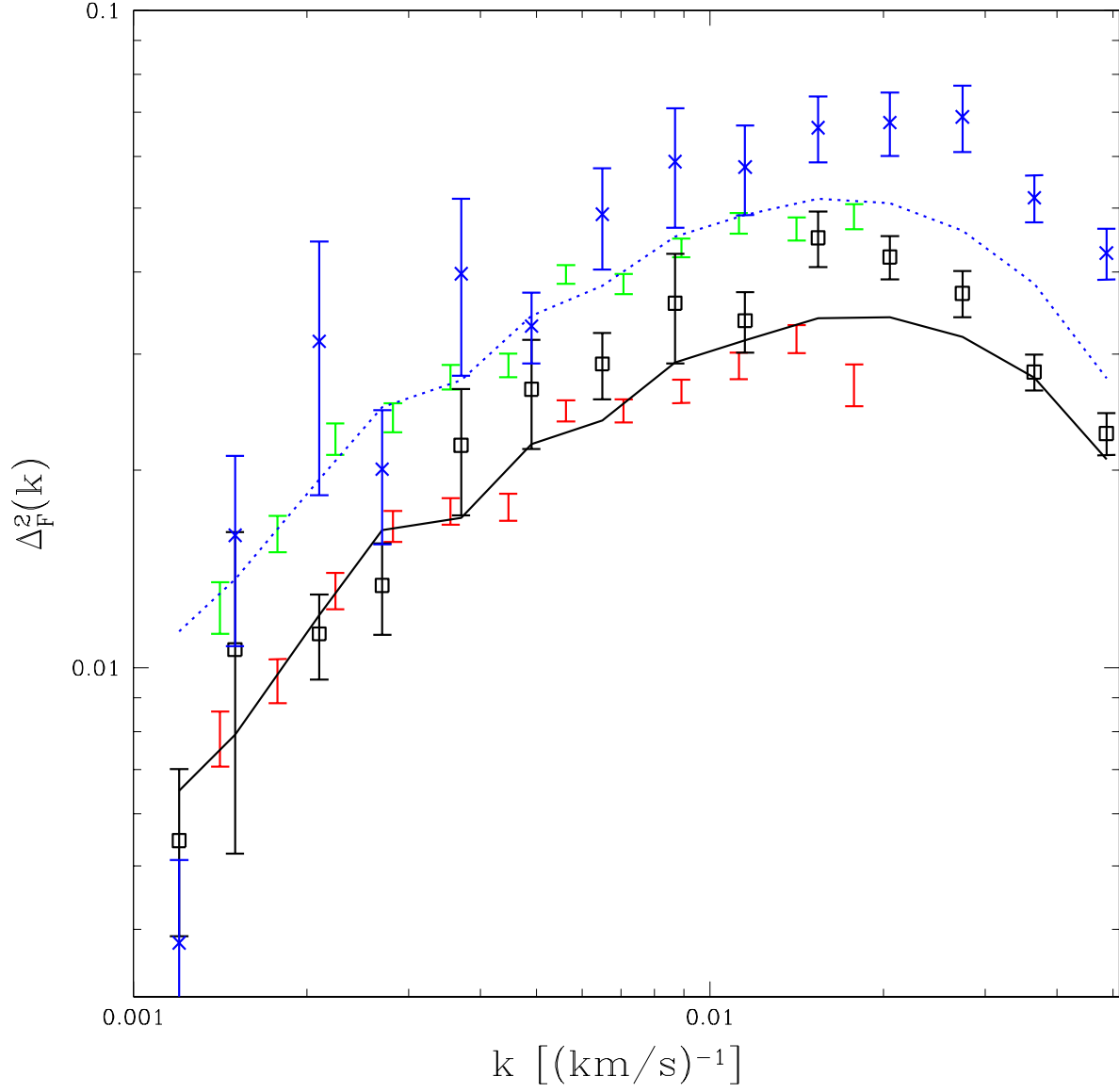


Fig. 37.— The black, solid line and open squares (blue, dotted line and crosses) show the fit prediction and measured points from Kim et al. (2004) at $z = 2.18$ (2.58). Red (green) error bars show our SDSS measurement at $z = 2.2$ (2.6).

$z = 2.58$ is 5.9σ below the prediction (as well as any reasonable extrapolation of the other Kim et al. (2004) points), and the points at increasingly high k are generally too high (at the highest k , this is a reflection of disagreement with the other high resolution data, but actually the agreement is not much better if we only include SDSS in the fit, because no model can fit the highest k SDSS points and then climb to match the higher k Kim et al. (2004) points). To reassure the reader that we are not playing games with the fitted curves, we also plot the SDSS points at $z = 2.2$ and 2.6 .

Since the Kim et al. (2004) results clearly have some problem, unless the other three measurements are all wrong (we will see that, with one exception, the other three agree with each other), we eliminate them from the rest of the comparison. A fit to SDSS, McDonald et al. (2000), and Croft et al. (2002) gives $\chi^2 = 269$ for ~ 210 dof (still a bad fit). Removing McDonald et al. (2000), reduces χ^2 by 44.3 (39 points), while removing Croft et al. (2002) reduces χ^2 by 97.2 (for 65 points this reduction would occur by chance only 0.6% of the time). Figure 38 shows the Croft et al. (2002) points, along with the fit prediction for them. The agreement is actually very good for 4 of the 5 redshift bins, while the $z = 2.47$ points are obviously out of place (these 13 points increase χ^2 by 54). Figure 39 shows the McDonald et al. (2000) points, along with the fit prediction for them (for this fit we removed the $z = 2.47$ Croft et al. (2002) points). The agreement is good, with the agreement at $z = 2.41$ disfavoring the anomalous Croft et al. (2002) $z = 2.47$ points; further investigation by R. Croft (private communication) does not reveal any obvious error in this redshift bin that would explain the anomaly. Note that the agreement of McDonald et al. (2000) and Croft et al. (2002) at high k adds weight to the idea that something is seriously wrong with Kim et al. (2004). Kim et al. (2004) show some comparisons with past results, and claim they agree, but these comparisons used custom redshift bins (i.e., not the bins in their table), and were not high precision (for example they compare the Croft et al. (2002) points at $z = 2.13$ to a bin with $z = 2.04$, so evolution cancels some of the amplitude offset, and they call the apparent $\sim 50\%$ difference at $k \sim 0.04(\text{km s}^{-1})^{-1}$ a “slight” excess).

5. Final Results Table and Directions for Use

Table 3 gives the primary power spectrum results. The columns are: z , the redshift of the bin; k , the wavenumber of the bin; $P_F(k, z)$, our final Ly α forest power spectrum result (along with the square roots of the diagonal elements of the error covariance matrix); P_{noise} , the noise power that was subtracted from each bin; and $P_{\text{background}}$, the background that was subtracted from each bin ($P_{1268,1380}$ adjusted according to the amount of noise in the forest, eqs. 3, 4, and 13). P_{noise} is just the noise subtracted from $P_{1041,1085}$ (a roughly comparable amount of noise was subtracted from the background, so to some degree these cancel in the final result). The table and the covariance matrix of the errors are available at <http://feynman.princeton.edu/~pmdonal/LyaF/sdss.html>. The covariance matrix must be used in any serious quantitative fitting. When using this table to constrain models, the following allowances should be made for residual systematic uncertainties:

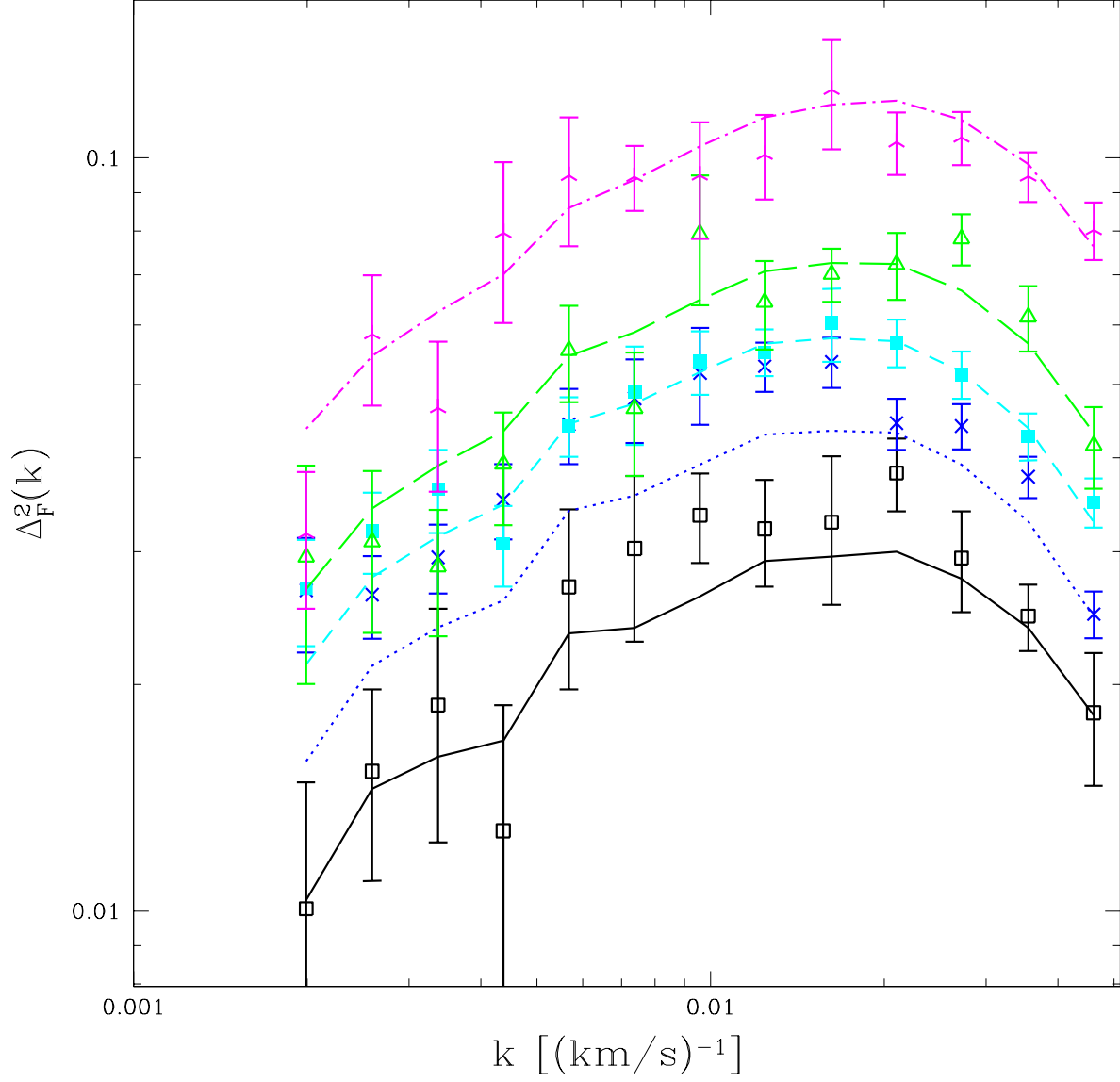


Fig. 38.— Measured points and fit prediction for the Croft et al. (2002) results. From bottom to top (roughly) — $z=2.13$: black, solid line, open square; $z=2.47$: blue, dotted line, 4-point star (cross); $z=2.74$: cyan, dashed line, filled square; $z=3.03$: green, long-dashed line, open triangle; $z=3.51$: magenta, dot-dashed line, 3-point star.

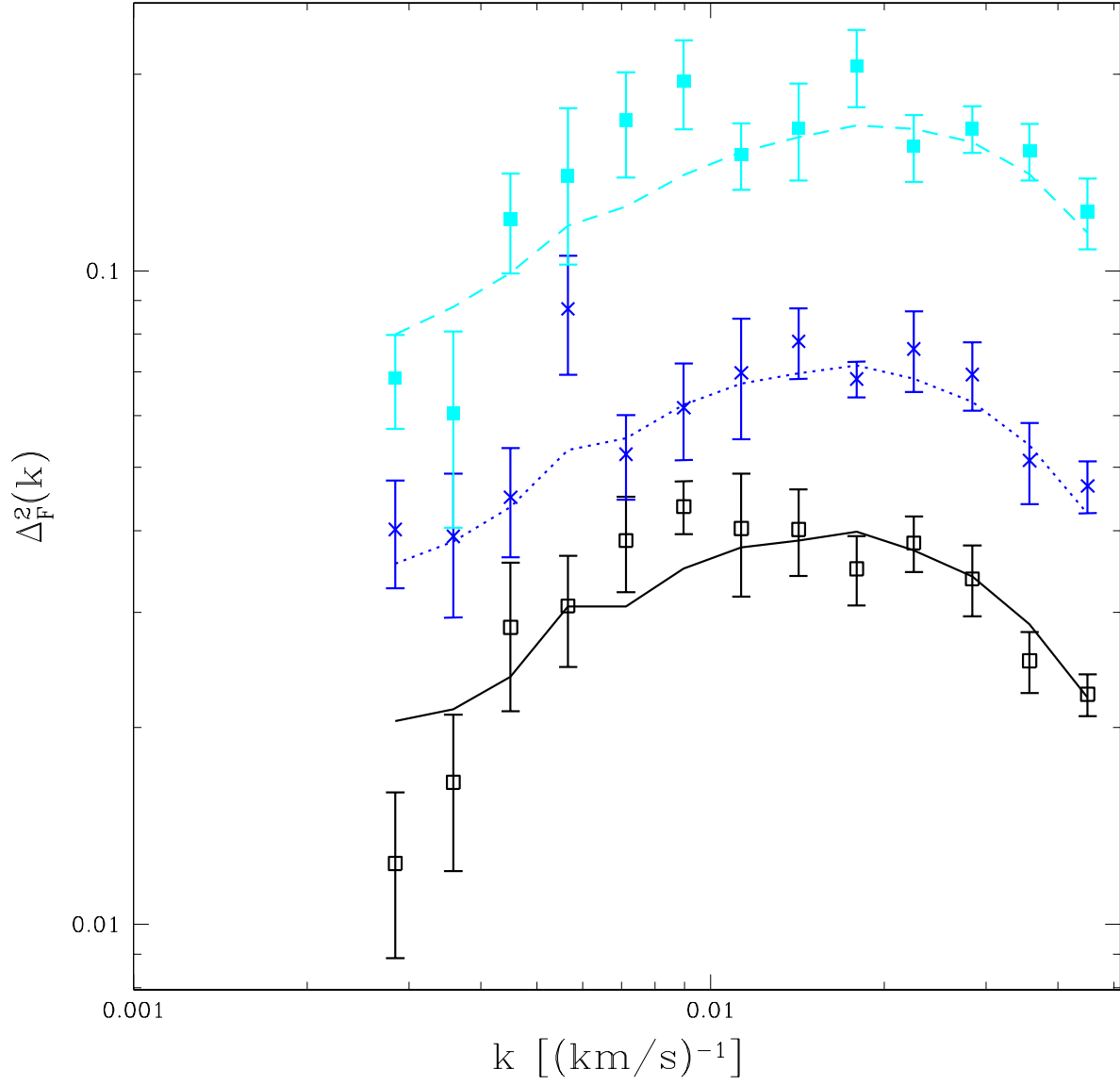


Fig. 39.— Measured points and fit prediction for the McDonald et al. (2000) results. From bottom to top (roughly) — $z=2.41$: black, solid line, open square; $z=3.00$: blue, dotted line, 4-point star (cross); $z=3.89$: cyan, dashed line, filled square;

- Allow $\pm 5\%$ rms error on the noise-power amplitude at each redshift. We do not have any reason to think the error is really this large, but, considering the complications related to the noise, we think it is prudent to include it. Operationally, we suggest subtracting $f_i P_{\text{noise}}(k, z_i)$ from $P_F(k, z_i)$, where f_i are free parameters in the fit (one for each redshift bin), and adding $\sum_i (f_i/0.05)^2$ to χ^2 .
- Allow $\pm (7 \text{ km s}^{-1})^2$ rms overall error on the resolution variance (i.e., the square of the rms width of the Gaussian resolution kernel). This is the expected size of the uncertainty due to flexure in the detector, although Figure 5 suggests that it may actually be smaller. Specifically, multiply $P_F(k, z)$ by $\exp(\alpha k^2)$, with α a free parameter in the fit, and add $[\alpha/(7 \text{ km s}^{-1})^2]^2$ to χ^2 .
- SiIII-Ly α cross-correlation must be accounted for. We have suggested a simple method – assume the cross-correlation has the same form as the Ly α -Ly α auto-correlation up to an amplitude that is a free parameter, and possibly include freedom in the correlation width and/or redshift evolution of the amplitude – but others could be devised (e.g., including SiIII in the simulated spectra through a parameterized semi-analytic model).

Table 3. $P_F(k, z)$ Results

z	k	$P_F(k, z)$	P_{noise}	$P_{\text{background}}$
2.2	0.00141	18.09 ± 1.74	6.20	3.29
2.2	0.00178	17.55 ± 1.34	6.07	2.71
2.2	0.00224	19.05 ± 1.21	6.20	2.46
2.2	0.00282	18.93 ± 1.03	6.23	2.45
2.2	0.00355	15.80 ± 0.74	5.92	1.97
2.2	0.00447	12.68 ± 0.60	5.57	1.18
2.2	0.00562	14.04 ± 0.53	5.85	1.04
2.2	0.00708	11.09 ± 0.45	5.87	1.17
2.2	0.00891	9.38 ± 0.39	6.06	1.20
2.2	0.01122	8.09 ± 0.38	6.87	1.40
2.2	0.01413	6.99 ± 0.34	8.48	1.32
2.2	0.01778	4.69 ± 0.35	10.52	0.87
2.4	0.00141	21.52 ± 1.91	5.70	3.72
2.4	0.00178	23.66 ± 2.09	5.65	2.63
2.4	0.00224	23.57 ± 1.39	5.65	2.45
2.4	0.00282	22.25 ± 1.24	5.58	2.47
2.4	0.00355	18.65 ± 0.89	5.31	1.82
2.4	0.00447	15.74 ± 0.65	5.15	1.18
2.4	0.00562	18.07 ± 0.68	5.43	0.79
2.4	0.00708	13.16 ± 0.51	5.32	1.07
2.4	0.00891	12.58 ± 0.41	5.71	1.14
2.4	0.01122	10.42 ± 0.41	6.27	1.29
2.4	0.01413	8.17 ± 0.36	7.32	1.22
2.4	0.01778	6.08 ± 0.33	9.37	0.91
2.6	0.00141	28.29 ± 2.55	6.78	4.02
2.6	0.00178	29.04 ± 1.85	6.68	3.18
2.6	0.00224	32.13 ± 1.76	6.76	2.65
2.6	0.00282	27.44 ± 1.39	6.63	2.41
2.6	0.00355	25.06 ± 1.09	6.52	1.79
2.6	0.00447	20.67 ± 0.85	6.40	1.27
2.6	0.00562	22.49 ± 0.72	6.69	1.05
2.6	0.00708	17.19 ± 0.60	6.71	1.17
2.6	0.00891	15.40 ± 0.51	7.12	1.05

Table 3—Continued

z	k	$P_F(k, z)$	P_{noise}	$P_{\text{background}}$
2.6	0.01122	13.25 ± 0.48	7.91	1.24
2.6	0.01413	10.25 ± 0.41	9.15	1.22
2.6	0.01778	8.43 ± 0.37	11.66	0.95
2.8	0.00141	37.25 ± 2.75	6.83	2.99
2.8	0.00178	37.52 ± 2.20	6.75	2.10
2.8	0.00224	38.74 ± 1.80	6.79	1.84
2.8	0.00282	37.12 ± 1.48	6.78	2.29
2.8	0.00355	30.11 ± 1.23	6.60	1.52
2.8	0.00447	25.67 ± 0.92	6.52	1.22
2.8	0.00562	25.74 ± 0.83	6.73	1.19
2.8	0.00708	22.54 ± 0.67	6.95	0.98
2.8	0.00891	20.12 ± 0.62	7.41	1.11
2.8	0.01122	15.89 ± 0.48	8.06	1.18
2.8	0.01413	13.04 ± 0.42	9.37	1.13
2.8	0.01778	9.63 ± 0.36	11.64	0.90
3.0	0.00141	46.36 ± 3.72	7.76	3.51
3.0	0.00178	42.53 ± 2.87	7.63	2.74
3.0	0.00224	47.66 ± 2.69	7.73	2.20
3.0	0.00282	42.20 ± 2.19	7.66	2.84
3.0	0.00355	36.99 ± 1.72	7.51	1.81
3.0	0.00447	29.47 ± 1.20	7.34	1.30
3.0	0.00562	30.12 ± 1.07	7.56	1.33
3.0	0.00708	24.30 ± 0.81	7.63	0.99
3.0	0.00891	22.51 ± 0.75	8.15	1.30
3.0	0.01122	18.75 ± 0.66	8.83	1.30
3.0	0.01413	14.33 ± 0.52	9.89	1.38
3.0	0.01778	11.26 ± 0.47	11.90	0.91
3.2	0.00141	54.73 ± 4.97	9.57	5.07
3.2	0.00178	49.72 ± 4.12	9.44	3.73
3.2	0.00224	52.86 ± 3.29	9.46	3.01
3.2	0.00282	48.44 ± 2.58	9.38	2.59
3.2	0.00355	44.01 ± 2.33	9.28	2.71
3.2	0.00447	35.12 ± 1.54	8.95	1.36

Table 3—Continued

z	k	$P_F(k, z)$	P_{noise}	$P_{\text{background}}$
3.2	0.00562	34.57 ± 1.41	9.15	1.28
3.2	0.00708	31.14 ± 1.19	9.40	1.29
3.2	0.00891	26.96 ± 0.95	9.80	1.48
3.2	0.01122	22.21 ± 0.83	10.49	1.80
3.2	0.01413	18.37 ± 0.70	11.74	1.38
3.2	0.01778	15.12 ± 0.66	14.07	1.18
3.4	0.00141	56.42 ± 5.85	11.12	4.08
3.4	0.00178	75.75 ± 5.33	11.32	2.41
3.4	0.00224	56.79 ± 3.87	10.97	2.09
3.4	0.00282	58.40 ± 3.43	11.04	3.16
3.4	0.00355	52.56 ± 2.85	10.96	2.62
3.4	0.00447	43.43 ± 2.21	10.76	2.00
3.4	0.00562	41.67 ± 1.73	10.99	2.04
3.4	0.00708	37.36 ± 1.43	11.28	1.55
3.4	0.00891	32.57 ± 1.19	11.87	1.77
3.4	0.01122	28.51 ± 1.15	13.06	2.00
3.4	0.01413	22.28 ± 0.88	14.63	1.63
3.4	0.01778	18.01 ± 0.79	17.80	1.23
3.6	0.00141	79.46 ± 8.33	15.11	2.25
3.6	0.00178	85.12 ± 8.28	14.90	2.87
3.6	0.00224	75.03 ± 5.88	14.87	3.28
3.6	0.00282	66.15 ± 4.98	14.51	3.30
3.6	0.00355	66.32 ± 4.08	14.59	2.26
3.6	0.00447	55.66 ± 3.36	14.22	1.33
3.6	0.00562	49.51 ± 2.72	14.17	1.28
3.6	0.00708	43.77 ± 2.15	14.41	1.62
3.6	0.00891	40.20 ± 1.93	15.09	1.98
3.6	0.01122	32.04 ± 1.63	15.72	1.61
3.6	0.01413	25.82 ± 1.31	17.26	1.25
3.6	0.01778	21.49 ± 1.23	21.11	1.51
3.8	0.00141	118.61 ± 15.47	22.58	6.89
3.8	0.00178	61.52 ± 9.80	20.93	6.40
3.8	0.00224	77.29 ± 7.09	20.91	5.07

Table 3—Continued

z	k	$P_F(k, z)$	P_{noise}	$P_{\text{background}}$
3.8	0.00282	71.78 ± 7.42	20.76	2.20
3.8	0.00355	77.49 ± 5.65	21.00	2.51
3.8	0.00447	59.12 ± 4.21	20.37	2.37
3.8	0.00562	57.53 ± 3.72	20.78	2.61
3.8	0.00708	56.25 ± 3.45	21.63	2.79
3.8	0.00891	42.46 ± 2.34	21.69	2.57
3.8	0.01122	36.93 ± 2.25	23.43	2.69
3.8	0.01413	29.52 ± 2.25	26.14	2.47
3.8	0.01778	27.72 ± 1.85	33.51	2.15

Note. — k has units $(\text{km s}^{-1})^{-1}$, power spectra have units km s^{-1} . The error covariance matrix must be used for any quantitative fitting.

6. Conclusions

We have analyzed a sample of 3035 quasar spectra measured by SDSS and covering the redshift range $2 < z < 4$. This data set is almost two orders of magnitude larger than previously available data sets. We have focused on the flux power spectrum in the redshift range $2.1 < z < 3.9$ and for modes $0.0013 \text{ (km s}^{-1}\text{)}^{-1} < k < 0.02 \text{ (km s}^{-1}\text{)}^{-1}$. The extraordinary size of the data sample leads to an order of magnitude reduction in errors compared to previous analyses. Consequently, to do justice to this data set requires a much more careful analysis than was needed in the past. To this end we have developed a new analysis pipeline using quadratic power spectrum estimation with near optimal performance. We applied this analysis to realistic mock spectra and demonstrated (after several tweaks) that the method performs as expected. We emphasize that realistic mock spectra are essential if one is to trust the results at the level of precision allowed by this data set. Our error estimation is based on bootstrap resampling, which works well here because the individual quasars are independent of each other. The errors were tested against mock spectra and found to be accurate. We also compared the bootstrap errors to Gaussian errors, finding them to be in general less than 20% higher than Gaussian.

Given the small errors on the recovered flux power spectrum the required control of systematic effects must be improved correspondingly as well. A significant part of this paper is devoted to this issue. We find several sources of contamination present in the data and develop methods to remove them. Metal absorption for metals with rest wavelength transitions above $\sim 1300\text{\AA}$, uncertainties in sky subtraction, and calibration errors can be subtracted essentially exactly by measuring the power on the red side of Ly α forest using the same observed wavelength range. We search for a contribution from metals with transitions close to Ly- α using a correlation function analysis, assuming that they are correlated with hydrogen. We find clear evidence of SiIII contamination and develop a simple and effective scheme to remove it. This procedure improves the χ^2 of the fit from 194 to 129 for ~ 104 degrees of freedom and is thus necessary for a satisfactory fit. We find no evidence of any other metal line contribution to the background subtracted flux power spectrum.

We reduce any contribution of the continuum to the flux power spectrum by dividing each spectrum by the mean quasar continuum. If contributions from quasar-to-quasar continuum differences are similar in different regions of the spectrum, then our subtraction of power from the red side of Ly α , as described above, should remove them. Several tests suggest that any residual contributions from continuum fluctuations are negligible. First, we measure the power in the mean continuum in several rest frame regions, finding it to be always small relative to our error bars, so power in quasar-to-quasar fluctuations has to be larger than power in the mean continuum itself to be significant. Second, measurements of the background in several rest frame regions place upper limits on the fluctuations in those regions. Third, a split of the Ly α forest region into two halves reveals no evidence that residual continuum fluctuations differ from one half to the other. Finally, estimating the continuum quasar-by-quasar using a principal component analysis does not change the power spectrum results significantly.

In section §4 we perform a series of tests to verify the robustness of the analysis against several modifications of the standard procedure and splits of the data. This reveals an interesting correlation between the power in the red side and the average noise (and some other properties of the spectra that correlate with noise, like the amplitude of the sky flux relative to the quasar flux). While we do not have a detailed explanation for this effect, we are able to remove it by modifying the standard procedure to include this correlation. From our full battery of tests, we conclude that systematic effects in the power spectrum measurement are not likely to significantly affect the results of cosmological fitting (i.e., it is likely that some effects remain formally significant relative to the errors on $P_F(k, z)$, but the shape of these systematic errors does not seem to correspond to a change in the cosmological parameters). This conclusion is further confirmed by the analysis of different subsets, which do not reveal any systematic deviations from those expected statistically.

In this paper we limit ourselves to the analysis of the flux power spectrum, without attempting to compare it to cosmological models. The results of this paper should thus be fairly noncontroversial and can be used by others who wish to perform their own cosmological analysis. Our own analysis will be presented in a separate publication, as will the cosmological implications that follow from it. We note that the expected error on the linear rms amplitude of fluctuations is $\sim 5\%$ and on the slope is ~ 0.024 , both at the pivot point $k = 0.009 \text{ (km s}^{-1}\text{)}^{-1}$. This should be compared to 10% error on the amplitude and 0.04 on the slope from the WMAP data at $k = 0.05 \text{ Mpc}^{-1}$ (Spergel et al. 2003). This data set provides very tight constraints on the amplitude and slope of the matter power spectrum. Many additional analyses can be performed using this data set, among them the mean flux evolution, cross-correlations between close pairs, and a bispectrum measurement. These will provide a wealth of additional information both on cosmology and on the state of the intergalactic medium at $2 < z < 4$, and they will allow us to test the basic picture of the Ly α forest that has emerged over the last decade. We believe that the unprecedented size of this data set will revolutionize our understanding of the high redshift universe; this work is merely a first step in this endeavor.

Funding for the creation and distribution of the SDSS Archive has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the U.S. Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS Web site is <http://www.sdss.org/>.

The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The Participating Institutions are The University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, University of Pittsburgh, Princeton University, the United States Naval Observatory, and the University of Washington.

Some of the computations used facilities at Princeton provided in part by NSF grant AST-0216105, and some computations were performed at NCSA.

Transition wavelengths used in this paper are from the Atomic Line List, <http://www.pa.uky.edu/~peter/atomic>

JS is grateful for support from the W.M. Keck Foundation and NSF grant PHY-0070928. DPS is supported by NSF grant AST03-07582.

REFERENCES

- Abazajian, K., et al. 2003, *AJ*, 126, 2081
- Abazajian, K., et al. 2004, *astro-ph/0403325*
- Bernardi, M., et al. 2003, *AJ*, 125, 32
- Bi, H. G., Boerner, G., & Chu, Y. 1992, *A&A*, 266, 1
- Blanton, M. R., Lin, H., Lupton, R. H., Maley, F. M., Young, N., Zehavi, I., & Loveday, J. 2003, *AJ*, 125, 2276
- Cen, R., Miralda-Escude, J., Ostriker, J. P., & Rauch, M. 1994, *ApJ*, 437, L9
- Croft, R. A. C., Weinberg, D. H., Bolte, M., Burles, S., Hernquist, L., Katz, N., Kirkman, D., & Tytler, D. 2002, *ApJ*, 581, 20
- Croft, R. A. C., Weinberg, D. H., Katz, N., & Hernquist, L. 1998, *ApJ*, 495, 44
- Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., & Schneider, D. P. 1996, *AJ*, 111, 1748
- Gnedin, N. Y., & Hamilton, A. J. S. 2002, *MNRAS*, 334, 107
- Gnedin, N. Y., & Hui, L. 1998, *MNRAS*, 296, 44
- Gunn, J. E., et al. 1998, *AJ*, 116, 3040
- Hall, P. B., et al. 2002, *ApJS*, 141, 267
- Hernquist, L., Katz, N., Weinberg, D. H., & Jordi, M. 1996, *ApJ*, 457, L51
- Hogg, D. W., Finkbeiner, D. P., Schlegel, D. J., & Gunn, J. E. 2001, *AJ*, 122, 2129
- Kim, T. ., Viel, M., Haehnelt, M. G., Carswell, R. F., & Cristiani, S. 2003, *ArXiv Astrophysics e-prints*
- Kim, T.-S., Viel, M., Haehnelt, M. G., Carswell, R. F., & Cristiani, S. 2004, *MNRAS*, 347, 355
- Lynds, R. 1971, *ApJ*, 164, L73
- Mandelbaum, R., McDonald, P., Seljak, U., & Cen, R. 2003, *MNRAS*, 344, 776
- McDonald, P., Miralda-Escudé, J., Rauch, M., Sargent, W. L. W., Barlow, T. A., & Cen, R. 2001, *ApJ*, 562, 52
- McDonald, P., Miralda-Escudé, J., Rauch, M., Sargent, W. L. W., Barlow, T. A., Cen, R., & Ostriker, J. P. 2000, *ApJ*, 543, 1

- Pier, J. R., Munn, J. A., Hindsley, R. B., Hennessy, G. S., Kent, S. M., Lupton, R. H., & Ivezić, Ž. 2003, *AJ*, 125, 1559
- Richards, G. T., et al. 2002, *AJ*, 123, 2945
- Ricotti, M., Gnedin, N. Y., & Shull, J. M. 2000, *ApJ*, 534, 41
- Schaye, J., Aguirre, A., Kim, T., Theuns, T., Rauch, M., & Sargent, W. L. W. 2003, *ApJ*, 596, 768
- Schaye, J., Theuns, T., Rauch, M., Efstathiou, G., & Sargent, W. L. W. 2000, *MNRAS*, 318, 817
- Seljak, U. . 1998, *ApJ*, 506, 64
- Seljak, U., McDonald, P., & Makarov, A. 2003, *MNRAS*, 342, L79
- Slanger, T. G., Cosby, P. C., Osterbrock, D. E., Stone, R. P. S., & Misch, A. A. 2003, *PASP*, 115, 869
- Smith, J. A., et al. 2002, *AJ*, 123, 2121
- Spergel, D. N., et al. 2003, *ApJS*, 148, 175
- Stoughton, C., et al. 2002, *AJ*, 123, 485
- Tegmark, M. 1997, *Phys. Rev. D*, 55, 5895
- Tegmark, M., Taylor, A. N., & Heavens, A. F. 1997, *ApJ*, 480, 22
- Theuns, T., Leonard, A., Efstathiou, G., Pearce, F. R., & Thomas, P. A. 1998, *MNRAS*, 301, 478
- Vogt, S. S., et al. 1994, in *Proc. SPIE Instrumentation in Astronomy VIII*, David L. Crawford; Eric R. Craine; Eds., Volume 2198, p. 362, Vol. 2198
- York, D. G., et al. 2000, *AJ*, 120, 1579
- Zaldarriaga, M., Scoccimarro, R., & Hui, L. 2001, *ApJ*, submitted (astro-ph/0111230)
- Zhang, Y., Anninos, P., & Norman, M. L. 1995, *ApJ*, 453, L57